

MISSISSIPPI LAW JOURNAL

ARTICLES

CONSTITUTIONAL LIMITS ON STATE TAXATION OF A
NONRESIDENT TRUSTEE: *GAVIN* MISINTERPRETS AND
MISAPPLIES BOTH *QUILL* AND *MCCULLOCH*
Joseph W. Blackburn

A PROLEGOMENA TO FEDERAL STATUTORY INTERPRETATION:
IDENTIFYING THE SOURCES OF INTERPRETIVE PROBLEMS
J. Gordon Christy

AMERICA'S ANTI-STANDING ARMY TRADITION AND THE
SEPARATE COMMUNITY DOCTRINE
Earl F. Martin

WORKING FOR FREE: IT OUGHT TO BE AGAINST THE (TAX) LAW
Richard Winchester

COMMENT

TILTING AT WINDMILLS: THE RELATIONSHIP BETWEEN MEN'S
NON-REVENUE SPORTS AND WOMEN'S SPORTS

RECENT DECISION

SECURITIES FRAUD—FEDERAL PREEMPTION—SLUSA PREEMPTS
STATE-LAW CLASS ACTIONS BROUGHT BY HOLDERS
OF COVERED SECURITIES

Introducing ... LexisNexis® Total Practice Solutions

- Client Development
- Research Solutions
- Practice Management
- Litigation Services

www.lexisnexis.com/lawschool

TOTAL PRACTICE SOLUTIONS

Client Development	Research Solutions	Practice Management	Litigation Services
---------------------------	---------------------------	----------------------------	----------------------------



LexisNexis®

Robert C. Khayat Scholarship Endowment

The *Mississippi Law Journal* established the endowment in February of 1995 to honor the dedicated service of Robert C. Khayat, former Law Journal member, professor and current University Chancellor. The purpose of the endowment is to assist distinguished men and women of the *Journal* in the pursuit of their legal education at the University of Mississippi. As of June 30, 2006, the corpus of the endowment was \$60,935.37. Each year deserving *Journal* members are awarded scholarships as selected by the Executive Board of the *Journal*. This year \$1,000 was awarded to Journal members. Please help us increase the corpus of the endowment so that we can increase the number and size of these scholarships.

Checks should be made payable to and mailed to:

The University of Mississippi Foundation
Post Office Box 8288
University, MS 38677

Please note on your check that your gift is designated for the Robert C. Khayat Scholarship Endowment.

Contributions are tax deductible.

The *Mississippi Law Journal* is published four times during the school year, including a Special Edition. Editorial and business offices: Post Office Box 849, University, Mississippi 38677.

Current subscription: \$35.00 per year. Single issue: \$12.00. Subscriptions are renewed automatically upon expiration unless the subscriber sends a notice of termination.

Change of address: Send address changes directly to the *Mississippi Law Journal*. Include name, new address (including zip code) and old address. Please notify the *Journal* forty-five days in advance to ensure prompt delivery. Unless a claim is made for non-receipt of the *Journal* within six months after the mailing date, the *Mississippi Law Journal* cannot be held responsible for supplying those issues without charge. Third class postage paid at University, Mississippi 38677 and additional mailing offices.

Manuscripts should be submitted electronically to **misslj@olemiss.edu**. contributors are requested to submit articles in 12-point type and double-spaced, in Microsoft Word, WordPerfect, or PDF. Citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (18th ed. 2005).

The *Mississippi Law Journal* is a member of the National Conference of Law Reviews.

To offer its contributors complete freedom, the *Journal* assumes no responsibility for the views expressed herein.

Back issues of the *Mississippi Law Journal* may be ordered directly from William S. Hein & Co., Inc., 1285 Main Street, Buffalo, New York 14029-1987. Orders may also be placed by calling William S. Hein & Co., Inc. at (800) 828-7571, via fax at (716) 883-8100 or email to order@wshein.com. Back issues are also available in PDF format through HeinOnline.



Digitized by the Internet Archive
in 2012 with funding from
LYRASIS Members and Sloan Foundation

Administrative Officers of the University of Mississippi

ROBERT C. KHAYAT	Chancellor
CAROLYN ELLIS STATON	Provost and Vice Chancellor for Academic Affairs
GLORIA KELLUM	Vice Chancellor for University Affairs
ALICE CLARK	Vice Chancellor for Research and Sponsored Programs
JOHNNY WILLIAMS	Vice Chancellor for Administration and Finance
WENDELL WEAKLEY	President/CEO, The University of Mississippi Foundation
TIMOTHY L. HALL	Associate Provost
THOMAS WALLACE	Vice Chancellor for Student Life
ANDREW P. MULLINS, JR.	Executive Assistant to the Chancellor
TIMOTHY L. WALSH	Associate Dean for Development
SAMUEL M. DAVIS	Dean of the School of Law, Director of Law Center, Jamie L. Whitten Chair of Law and Government and Professor of Law

THE LAW SCHOOL FACULTY

GUTHRIE T. ABBOTT,	<i>Professor Emeritus and Butler, Snow, O'Mara, Stevens and Cannada Lecturer in Law Emeritus</i>
RICHARD L. BARNES,	<i>Professor of Law and Leonard B. Melvin, Jr. Distinguished Lecturer in Law</i>
DEBORAH H. BELL,	<i>Professor of Law, Mississippi Defense Lawyers Association Distinguished Lecturer and Director of the Civil Legal Clinic</i>
KRISTI BOWMAN,	<i>Visiting Professor of Law</i>
JOHN R. BRADLEY,	<i>Professor of Law</i>
PHILLIP W. BROADHEAD,	<i>Director, Clinical Appeals Program and Clinical Professor of Law</i>
CHARLES BROWER, II,	<i>Croft Associate Professor of International Law and Jessie D. Puckett, Jr. Lecturer in Law</i>
MERCER E. BULLARD,	<i>Assistant Professor of Law</i>
LARRY BUSH,	<i>Professor Emeritus of Law</i>
WILLIAM M. CHAMPION,	<i>Professor Emeritus and Mississippi Defense Lawyers Association Distinguished Lecturer Emeritus</i>
THOMAS K. CLANCY,	<i>Director of the National Center for Justice and the Rule of Law and Visiting Professor of Law</i>
GEORGE C. COCHRAN,	<i>Professor of Law</i>
AARON S. CONDON,	<i>Professor Emeritus of Law</i>

JOHN M. CZARNETZKY, *Associate Professor of Law and Mitchell, McNutt, and Sams Lecturer*

DONNA DAVIS, *Associate Professor of Law*

JASON DERRICK, *Acting Assistant Professor of Legal Writing*

KYLE DUNCAN, *Assistant Professor of Law*

THOMAS R. ETHRIDGE, *Professor Emeritus of Law*

D. MICHAEL FEATHERSTONE, *Professor Emeritus of Law*

MOLLY FERGUSSON, *Acting Assistant Professor of Legal Writing*

JOANNE I. GABRYNOWICZ, *Director of the National Remote Sensing and Space Law Center and Research Professor of Law*

KRIS L. GILLILAND, *Director of Law Library and Associate Professor of Law*

CHRISTOPHER GREEN, *Assistant Professor of Law*

KAREN O. GREEN, *Professor of Law and Mississippi Defense Lawyers Association Distinguished Lecturer*

MATTHEW R. HALL, *Assistant Professor of Law*

TIMOTHY L. HALL, *Associate Provost and Professor of Law*

MARC HARROLD, *Counsel for National Programs and Visiting Professor of Law, National Center for Justice and the Rule of Law*

MICHAEL H. HOFFHEIMER, *Professor of Law and Mississippi Defense Lawyers Association Distinguished Lecturer*

ROBERT C. KHAYAT, *Chancellor and Professor of Law*

PATRICIA KRUEGER, *Acting Assistant Professor of Legal Writing and Director of Academic Excellence Program*

KATE MARGOLIS, *Visiting Professor of Law*

DON MASON, *Associate Director of the National Center for Justice and the Rule of Law and Lecturer*

KALI MURRAY, *Assistant Professor of Law*

GARY MYERS, *Professor of Law*

JACK W. NOWLIN, *Associate Professor of Law and Jessie D. Puckett, Jr., Lecturer*

E. FARISH PERCY, *Assistant Professor of Law*

LARRY J. PITTMAN, *Professor of Law and Jessie D. Puckett, Jr. Lecturer*

LISA S. ROY, *Assistant Professor of Law*

RONALD J. RYCHLAK, *Associate Dean for Academic Affairs, Professor of Law and Mississippi Defense Lawyers Association Lecturer*

PAUL M. SECUNDA, *Assistant Professor of Law*

JACQUELINE SERRAO, *Associate Director of the National Remote Sensing and Space Law Center*

HANS P. SINHA, *Director of Prosecutorial Externship Program and Clinical Professor of Law*

CAROLYN ELLIS STATON, *Provost, Vice Chancellor for Academic Affairs and Professor of Law*

ROBERT A. WEEMS, *Professor of Law and Butler, Snow, O'Mara, Stevens and Cannada Lecturer in Law*

C. JACKSON WILLIAMS, *Director of Legal Writing Program and Acting Assistant Professor of Legal Writing*

MISSISSIPPI LAW JOURNAL

PUBLISHED BY STUDENTS AT THE
UNIVERSITY OF MISSISSIPPI
SCHOOL OF LAW

VOLUME 76

FALL 2006

NUMBER 1

COPYRIGHT © 2006 MISSISSIPPI LAW JOURNAL, INC.

Editor-in-Chief
JOSHUA S. KYLE

Executive Editor
TRESA BARKSDALE

Articles Editors
JARED W. EASTLACK
MICHAEL J. GORMAN
NANCY WILLIAMS

Business Manager
KATIE HILL

Notes & Comments Editor
DAVID A. O'NEAL

Mississippi Cases Editor
KATIE E. WETHERBEE

ASSOCIATE EDITORS

Managing Editor
RYAN PRATT

Developmental Editor
NICHOLAS B. HOLTZ

Articles Editors
AMANDA BECKETT
JIM FARRELL
JOSEPH W. GILL
ELEANOR LAPORTE
KARA LINCOLN
KATHLEEN SHIELDS

Symposium Editor
KATE MAULDIN

Membership Director
MARY CLIFT HITT

Office Manager
JAKE DICKERSON

Mississippi Cases Editors
EMILY FERRIS
CAMPBELL MCLAURIN
SLATES VEAZEY

Notes & Comments Editors
LOREN HENAGAN
MAGGIE MAYNORD

Administrative Assistant
PEGGY C. NAIL

Faculty Advisors
DEBORAH H. BELL
JOHN CZARNETZKY
ROBERT A. WEEMS

STAFF

ELIZABETH ARCHER
CASEY BUTTS
ROBERT L. BOUIS, III
NATHAN BURROW
HUNTER CHANDLER
KATHRYN E. DENNIS
JONATHAN FESTA
JENNIFER FORMAN
TAYLOR HECK
ALEXANDRA HUTTON
MOLLY JEFFCOAT
JOSH JOHNSON
BRENT KLEIN
McCANN LEFEVE
WILLIAM A. LEWIS
COURTNEY LEYES
ERIC MCADAMIS

CAROLINE MCKIBBEN
TAYLOR MCNEEL
ALYSSON MILLS
SCOTT MURRAY
RYAN O'BEIRNE
ERIC PATTERSON
FRANK PELLEGRINO
RONNIE WADE ROBERTSON
LANE RICE
CLARE SMITH RUSH
A. NICOLE SEALE
NATALYA SEAY
REX M. SHANNON, III
JOSEPH TULLOS
AMANDA WALLIS
KIM WATSON
LANE WILLIAMSON

Mississippi Law Journal

Published by Students at the
University of Mississippi
School of Law

VOLUME 76

FALL 2006

NUMBER 1

Copyright © 2006 Mississippi Law Journal, Inc.

TABLE OF CONTENTS

ARTICLES

- CONSTITUTIONAL LIMITS ON STATE TAXATION OF A
NONRESIDENT TRUSTEE: GAVIN MISINTERPRETS AND
MISAPPLIES BOTH *QUILL* AND *McCULLOCH*
..... *Joseph W. Blackburn* 1

- A PROLEGOMENA TO FEDERAL STATUTORY INTERPRETATION:
IDENTIFYING THE SOURCES OF INTERPRETIVE
PROBLEMS
..... *J. Gordon Christy* 55

- AMERICA'S ANTI-STANDING ARMY TRADITION AND THE SEPARATE
COMMUNITY DOCTRINE
..... *Earl F. Martin* 135

- WORKING FOR FREE: IT OUGHT TO BE AGAINST THE (TAX) LAW
..... *Richard Winchester* 227

COMMENT

- TILTING AT WINDMILLS: THE RELATIONSHIP BETWEEN MEN'S
NON-REVENUE SPORTS AND WOMEN'S SPORTS
..... *Ronnie Wade Robertson* 297

RECENT DECISION

- SECURITIES FRAUD—FEDERAL PREEMPTION—SLUSA PREEMPTS
STATE-LAW CLASS ACTIONS BROUGHT BY HOLDERS OF COVERED
SECURITIES
..... *Rex M. Shannon III* 323

CONSTITUTIONAL LIMITS ON STATE TAXATION OF A NONRESIDENT TRUSTEE: GAVIN MISINTERPRETS AND MISAPPLIES BOTH *QUILL* AND *McCULLOCH*

*Joseph W. Blackburn**

I. INTRODUCTION

In *Quill Corp. v. North Dakota*,¹ the United States Supreme Court established substantive limitations to state taxation under the Fourteenth Amendment's Due Process Clause² and under the Commerce Clause's negative sweep.³ *Quill* is recognized as the principal and preeminent statement of these important Constitutional principles limiting state taxing jurisdiction and authority.⁴

Certainly, states often view restrictions on their taxing authority as unwelcome.⁵ A recent opinion of the Supreme

* J.D. University of Virginia. Palmer Professor of Taxation, Cumberland School of Law, Samford University.

¹ *Quill Corp. v. North Dakota*, 504 U. S. 298 (1992).

² U.S. CONST. amend. XIV, § 1.

³ *Quill*, 504 U.S. at 309. See U.S. CONST. art. I, § 8, cl. 3; *Gibbons v. Ogden*, 22 U.S. (9 Wheat.) 1 (1824) (recognizing the negative power of the Commerce Clause).

⁴ See, e.g., *Scholastic Book Club, Inc. v. Dep't of Treasury*, 567 N.W.2d 692 (Mich. Ct. App. 1997); *Magnetek Controls, Inc. v. Dep't of Treasury*, 562 N.W.2d 219 (Mich. Ct. App. 1997); *Town Crier, Inc. v. Dep't of Revenue*, 733 N.E.2d 780 (Ill. App. Ct. 2000).

⁵ See, e.g., *Blue v. Dep't of Treasury*, 462 N.W.2d 762 (Mich. Ct. App. 1990); John Caher, *State Tax Ruling Roils Controversy with Neighbors*, 226 N.Y.L.J. 1, (2001) (analyzing a fight between New York and Connecticut over allocation of a Connecticut resident's salary from a New York University). See Bernard E. Jacob, *An Extended Presence, Interstate Style: First Notes on a Theme from Saenz*, 30 HOFSTRA L. REV. 1133 (2002).

Court of Connecticut represents the most extreme of these cases. In *Chase Manhattan Bank v. Gavin*,⁶ the Supreme Court of Connecticut⁷ upheld the imposition of Connecticut income taxes on a New York trustee.⁸ The trustee had no presence whatsoever in Connecticut.⁹ The trustee, all trust assets, applicable state trust law, and all trust administration were outside of Connecticut.¹⁰ Indeed, the essential "critical link" cited by the Supreme Court of Connecticut to justify Connecticut's taxation of the New York *inter vivos* trustee's income was that the trust's sole, noncontingent beneficiary was a Connecticut resident.¹¹ Connecticut's sole contact with the testamentary trusts was the decedent settlor's residence in Connecticut at the time of death, in one instance sixty-three years earlier.¹² Such testamentary trusts, however, were formed under New York law with New York administration and New York Trustee.¹³ In such circumstances, domicile of the decedent should not be controlling.¹⁴

The Connecticut tax at issue was imposed upon the New York trustee's entire accumulated, undistributed income and gains from the trust property.¹⁵ Connecticut's authority to tax its own resident, *i.e.*, the *inter vivos* trust's Connecticut beneficiary, on the beneficiary's income from any domestic or foreign

⁶ *Chase Manhattan Bank v. Gavin*, 733 A.2d 782 (Conn. 1999) [hereinafter *Gavin*].

⁷ Concurring in Justice Borden's opinion were Justices Callahan, Norcott, Palmer, and Peters. Justices McDonald and Berdon dissented. *Id.* at 806.

⁸ The case involved taxation of four testamentary trusts of Connecticut decedents and one *inter vivos* trust (the Adolffson Trust) established by a Connecticut resident as settlor. The testamentary trusts ranged from 24 to 63 years old, having been formed many years previously under New York law by Connecticut decedents. *Id.* at 786-87.

⁹ *Id.* at 787.

¹⁰ *Id.*

¹¹ *Id.* at 802.

¹² *Id.* at 787.

¹³ See *infra* note 142 (discussing *District of Columbia v. Chase Manhattan Bank*, 689 A.2d 539 (D.C. 1997)).

¹⁴ *Id.* Domicile should not be controlling, or even relevant. *Id.*

¹⁵ *Gavin*, 733 A.2d at 785, 802.

source is unquestioned.¹⁶ This would clearly include distributed and even distributable income of a foreign trustee *if* Connecticut law had imposed such a tax on the resident beneficiary.¹⁷

Gavin, however, held the New York trustee's undistributed or undistributable income taxable. In doing so, the Connecticut Supreme Court failed to honor the admonitions of *Quill* that a taxpayer must initiate its own due process contacts with the taxing jurisdiction.¹⁸ The *Gavin* opinion also gave little attention to the Supreme Court's Commerce Clause analysis¹⁹ which was essential to the Court's rejection of North Dakota's tax levy against *Quill*. To the limited extent *Gavin* addressed the Connecticut tax's discriminatory effect on interstate commerce, it misapplied *Quill*.²⁰ The Connecticut justices also refused to follow a clear line of well-reasoned and authoritative cases²¹ that went clearly against their *Gavin* holding.²²

The opinion then played a shell game of words in order to artificially place its facts as to the *inter vivos* trust within those of a single, California case it cited as precedent,²³ the facts, law, and holding of which were completely different.²⁴ The *McCulloch* case involved a unique California statute which was the predecessor of federal throwback rules.²⁵

The purpose of this article is to point out flaws in *Gavin's* analysis and thereby hopefully lessen any inappropriate mischiefs to Constitutional principles that may result from its

¹⁶ See *Schaffer v. Carter*, 252 U.S. 37 (1920).

¹⁷ *McGuire v. Trefry*, 253 U.S. 12 (1920). See discussion *infra* Part III-A.

¹⁸ See *supra* notes 1-4 and accompanying text; see also *infra* Parts V, VI, and VII.

¹⁹ See discussion *infra* Part VII.

²⁰ See *infra* note 221 and accompanying text.

²¹ See *Nenno & Sparks, Delaware Dynasty Trust*, at 32, published by Wilmington Trust, Wilmington Delaware: "The court's constitutional analysis particularly for the *inter vivos* trust is unpersuasive."

²² See *infra* note 316.

²³ *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412 (Cal. 1964) (citations omitted), *appeal dismissed*, 379 U.S. 133, 85 S. Ct. 278 (1964).

²⁴ See discussion *infra* Part VIII.

²⁵ See *infra* note 264.

misguided holding.²⁶

The article first analyzes pre-*Quill* case law, and then discusses the *Quill* opinion. Thereafter, *Gavin's* holding will be reviewed in light of this line of precedent cases, including its clearly misplaced reliance on *McCulloch*,²⁷ *the only precedent it cited for taxation of the inter vivos trust*.

It is the recurring theme of this article that Constitutional due process analysis of state taxation of trusts has lost its way. Courts must return to the fundamental principles of the law of trusts. A "trust" is not "a legal person" or "an entity." In applying due process minimal contacts analysis, one must first determine what a "trust" is and what person's contacts with the taxing state must be analyzed. That is, in evaluating sufficiency of nexus for due process purposes, the law looks to the contacts of the "trustee" as a legal entity, not to contacts of a "trust", an abstraction, or to a beneficiary?

II. SYNOPSIS OF THE LAW OF TRUSTS: *INTER VIVOS V. TESTAMENTARY*

A. A "trust" is not a legal person under state law.

The artificial nature of trusts complicates jurisdictional analysis. A trustee is a legal person whether corporate or human. The term "trust" means "a *fiduciary relationship* with respect to property, subjecting the person [trustee] by whom the title to the property is held to equitable duties to deal with the property for the benefit of another person [beneficiary], which arises as a result of a manifestation of an intention to create it."²⁸ Thus a trust is not a legal entity or person, but merely a "fiduciary relationship" whereby the trustee, a legal person, owns property subject to the duty to administer it on

²⁶ See Jordan M. Goodman, *State Taxation of Business Trusts: Limits, Concerns and Opportunities*, 9 Feb. J. MULTISTATE TAX'N 6 (2000).

²⁷ See *supra* note 23.

²⁸ RESTATEMENT (SECOND) OF TRUSTS, §2 (1959) (emphasis added). I SCOTT, TRUSTS § 2.3 (4th ed. 1987); see also *Randolph Found. v. Appeal from Probate Court*, No. XO5CV980167903S, 2001 WL 418059, at *16 (Conn. Super. Ct. Apr. 3, 2001) (trust is an abstraction).

behalf of the beneficiary, a separate legal person. The property and associated income subject to this fiduciary duty is the fiduciary's, i.e., the trustee's. Such property and affiliated income is not held by any separate, distinct legal entity referred to as "the trust," and not by the beneficiary.²⁹ A trust is a fiction, i.e. an abstraction.³⁰

As an "abstraction," a trust may not sue or be sued or hold and transfer title to property in its own name.³¹ A trustee can sue or be sued based on the trustee's actions or to determine issues pertaining to property held in such trustee's name.³² The trustee need not sue in his name as trustee but solely in the trustee's own name without reference to the trust relationship.³³ Furthermore, a beneficiary is in no way an agent of the trustee nor a necessary or even proper party to an action by the trustee.³⁴ A trust has no existence separate from the persons who hold differing interests therein. Modern "entity law"³⁵ generally recognizes the following legal persons: corporations, legal liability companies, general partnerships, limited liability partnerships, limited partnerships, REIT'S, and qualifying business trusts.³⁶

The law of trusts applies to both *inter vivos* trusts and testamentary trusts. A trust may be created during the owner's lifetime through transfer of property by deed to a trustee. For testamentary trusts, property is merely transferred from the decedent by will to a trustee. Transfer of prop-

²⁹ In the 1913 Revenue Act, Section II, Paragraph D, the fiduciary withheld tax for the beneficiary's tax liability. Since 1916 taxes imposed on a "trust" are in reality fiduciary income taxes imposed on the fiduciary person in such person's capacity as fiduciary. See, e.g., 26 U.S.C. §641(b) (2000); *McCauley v. Comm.*, 44 F.2d 919 (5th Cir. 1930).

³⁰ See *Randolph Foundation*, *supra* note 28.

³¹ See *infra* note 32.

³² "The trustee can maintain such actions at law or suits in equity or other proceedings against a third person as he could maintain if he held the trust property free of trust." RESTATEMENT (SECOND) OF TRUSTS, §280.

³³ *Id.* at §280(h).

³⁴ *Id.* at §280(i).

³⁵ See Philip I. Blumberg, *The Corporate Entity In An Era of Multinational Corporations*, 15 DEL. J. CORP. L. 283, 286 (1990).

³⁶ *Current Issues in Conveyance Practices*, 15438 NBI-CLE 1, 7.

erty to a trustee by deed or by will merely reflects two different methods of creating a trust relationship.³⁷ The method of transfer does not alter in any way requirements for creation of a trust relationship, e.g. settlor or testator's capacity, intention to create a trust, designation of a beneficiary, duties to be performed by a trustee, etc.³⁸ Each method requires an effective transfer of property to a trustee, irrespective of the instrument used to accomplish such a transfer, i.e. deed, assignment, or will.

The law of the settlor's or testator's residence generally³⁹ controls the effectiveness of the transferring instrument, whether that instrument is a deed or a will.⁴⁰ In the context of an *inter vivos* transfer, the transferor's compliance with a state's statute of frauds or formalities for real estate deeds allows the resident transferor to effectively transfer property to the trustee.⁴¹ In the context of a testamentary transfer, transferor/decedent's compliance with a state's statute of wills and probate formalities similarly allows the resident to effectively transfer property to the trustee by will at death.⁴²

These resident state laws directly apply to and benefit the state's resident, whether *inter vivos* transferor or decedent. These and other state laws of the transferor's domicile indirectly benefit any transferee, resident or nonresident, on the other side of any such transfer or transaction.⁴³

Just as with an *inter vivos* trust, a testator may designate in his will which state law, other than the law of testator's residence, to apply to governance of the trust.⁴⁴ Likewise, the

³⁷ ULA Trust Code § 401: "Methods of Creating Trust. A trust may be created by: (1) transfer of property to another person as trustee during the settlor's lifetime or by will or other disposition taking effect upon the settlor's death." See also RESTATEMENT (SECOND) OF TRUSTS § 17.

³⁸ RESTATEMENT (SECOND) OF TRUSTS § 402.

³⁹ Except, for example, effective transfer of real estate is determined by the law of the situs state.

⁴⁰ See *infra* note 46.

⁴¹ In the context of sales, compliance with the U.C.C. or a state's common law of contracts similarly enables a transferor to effectively transfer property.

⁴² See *infra* note 46.

⁴³ See discussion of *Quill*, *infra* notes 192-94 and accompanying text.

⁴⁴ ULA Trust Code § 107: "Governing Law. The meaning and effect of a trust

testator may designate administration in a non-domiciliary state and may select a nonresident trustee.⁴⁵ In contrast, the validity of the testator's will as an effective dispositive instrument is determined by the law of the testator's state of residence.⁴⁶ Likewise, whether an instrument effectively conveys an interest in land is determined by the law of the situs of such land.⁴⁷

State law benefits afforded resident persons cannot, under due process, be attributed as an indirect benefit to every nonresident with whom such residents have personal or commercial relationships or dealings.⁴⁸ If one state can impose tax burdens on the basis of such imputed state law benefits, every state, county, municipality or other taxing jurisdiction will have wholly unfettered ability to tax nationwide. Such attribution, reattribution, or re-reattribution⁴⁹ would create an

are determined by: (1) the law of the jurisdiction *designated in the terms* unless the designation of that jurisdiction's law is contrary to a strong public policy of the jurisdiction having the ost significant relationship to the matter at issue; or ..." (emphasis added).

⁴⁵ ULA Trust Code § 108: "Principal Place of Administration. (a) Without precluding other means for establishing a sufficient connection with the designated jurisdiction, terms of a trust *designating* the principal place of administration are valid and controlling if: (1) a trustee's principal place of business is located in or a trustee is a resident of the designated jurisdiction; or (2) all or part of the administration occurs in the designated jurisdiction" (emphasis added).

⁴⁶ WILLIAM A. FRATCHER, *SCOTT ON TRUSTS* §§ 649-650 (4th ed. 1989); *see also* *Hanson v. Denckla*, 357 U.S. 235 (1958).

⁴⁷ FRATCHER, *supra* note 46, §§ 649-50.

⁴⁸ *See infra* note 228 and accompanying text.

⁴⁹ In a recent, well-reasoned opinion, the Alabama Administrative Law Division (the "Division") rejected an attempt by the Alabama Department of Revenue ("ADOR") to impute such benefits under Alabama law. ADOR was attempting to impute activities of an Alabama limited partnership to its nonresident limited partner. Lanzi, the limited partner, did not participate in management of the partnership or its assets, did not own property, earn income, or conduct business in Alabama, and resided in Georgia. The Division's opinion held that the limited partnership was treated as a separate legal entity under Alabama law. The opinion then stated, "[b]ecause a partnership is a separate entity under current Alabama law, the presence and activities of a partnership in Alabama *cannot be attributed to its nonresident [limited] partners* for nexus purposes. Separate legal entities must be recognized as such." *Lanzi v. Ala. Dep't of Revenue*, No. INC. 02721 (Ala. Admin. Law Div. Sept. 26, 2003) (emphasis added). This author acknowledges that he, along with Donald E. Johnson and David M. Wooldridge of

infinite chain unwittingly and wrongly subjecting every person to the jurisdiction of every political subdivision in the nation. That is why *Quill's* due process analysis requires that a non-resident knowingly and deliberately direct its contacts at the foreign jurisdiction.⁵⁰

B. What legal "person"⁵¹ must have "purposefully directed its activities"⁵² at a state in order to meet the minimum contacts standard for due process purposes?

Sufficiency of contacts depends on the nature of the jurisdiction in question. For an *in rem* action, contacts with the subject property must be determined.⁵³ For *in personam* jurisdiction over a "person," the contacts of such person must be determined. Analysis of jurisdiction to tax involves both *in rem* and *in personam* jurisdiction. If a state has jurisdiction as to property, then it has jurisdiction to tax the income from such property. That does not, however, equate to *in personam* jurisdiction over the owner of such property.⁵⁴

For example, a single member limited liability company ("LLC") is a disregarded entity⁵⁵ for income tax purposes. However, under state laws it is a legal entity which can hold title to property, conduct business in its own name, and sue or be sued in its own name.⁵⁶ Thus income tax laws artificially

Sirote & Permutt, P.C., represented Mr. Lanzi. The case is presently on appeal to the Alabama Court of Civil Appeals—the Circuit Court of Montgomery County having held that Alabama did have jurisdiction to tax the nonresident.

⁵⁰ See *infra* notes 192-94 and accompanying text; see also *Burger King v. Rudzewicz*, 471 U.S. 462 (1985) (taxpayer purposefully directed activities at the taxing state).

⁵¹ The term "person" in this article may be an individual, i.e., a human person, or a legally recognized entity such as a corporation, limited liability company, or partnership.

⁵² See *infra* note 192.

⁵³ Bogert, *The Law of Trusts and Trustees*, §292.

⁵⁴ Lowy and Vasquez, "When is it Unconstitutional for States to Tax Nonresident Members of Limited Liability Companies," [VOL NO.] STATE TAX NOTES 633 (May 2003). ("The Power to Tax is empty without the power to collect the tax". Jurisdiction to tax income is interdependent with personal jurisdiction over the taxpayer).

⁵⁵ See, e.g., I.R.S. Chief Couns. Advisory, CCA 200216028 (Apr. 18, 2002).

⁵⁶ *Id.*; For a general discussion, see CARTER G. BISHOP & DANIEL S.

render such an entity as nonexistent for income tax purposes. For income tax purposes, the person who is the LLC's single member is directly taxed.

Assume such an LLC is formed under the laws of State X, actively conducts business in State X, and owns property in State X. State X disregards the LLC for income tax purposes, thereby taxing Y, its single member. If Y is a nonresident of State X, having no direct contact with State X, can State X tax nonresident Y based solely on Y's ownership of the single member interest in the LLC?⁵⁷

Artificial, constantly evolving tax classifications and elections, should not and do not⁵⁸ control the parameters of constitutional due process for tax or other jurisdictional purposes. States have broad powers to impose income taxes subject only to constitutional limitations and their own states laws.⁵⁹ For example, in the context of a trust, tax on income may and at times is imposed on the settlor/grantor, the trustee, and/or the beneficiary. In the context of *Gavin*, Connecticut could tax the Connecticut resident settlor/grantor of a New York *inter vivos* trust. Taxation could be based on grantor trust rules⁶⁰ or on deemed realization of gain upon transfer of the Connecticut property of such a resident to the out-of-state trustee.⁶¹ Sub-

KLEINBERGER, LIMITED LIABILITY COMPANIES: TAX AND BUSINESS LAW, *available at* 2000 WL 12016 (2000).

⁵⁷ What if this entity under "check the box" Treas. Reg. § 301.7701 (2006), had elected to be taxed as a corporation? Should the vagaries of a tax election determine due process nexus? What if the LLC elected to be taxed as a corporation, but then made a subchapter S election? Should due process analysis change again?

⁵⁸ See *supra* note 49. If *Quill* had been a subchapter S corporation, would its shareholder owners then have nexus with North Dakota, individually, i.e., would the corporation's contacts be attributed to its shareholders merely because of a tax election?

⁵⁹ In addition to due process and commerce clause restrictions, "realization" is a further constitutional constraint imposed on income taxation. *Comm'r v. Glenshaw Glass Co.*, 349 U.S. 925 (1955).

⁶⁰ I.R.C. §§ 671-679.

⁶¹ See, e.g., I.R.C. §§ 644, 877, regarding tax on transfer of appreciated property. Also, grantor trust rules should apply to all "foreign", i.e. out of state trusts. See, e.g., I.R.C. § 679.

ject to realization issues,⁶² Connecticut could have chosen to tax the Connecticut resident beneficiary on income earned from property or activities of the nonresident trustee, just as California had done.⁶³ But, if Connecticut attempts to actually tax the nonresident trustee, the trustee itself must have knowingly directed its contacts at Connecticut.⁶⁴

III. PRE-QUILL CASES

Predictably, there is a long line of state and federal cases addressing state jurisdiction to tax foreign trustees within the limits of federal due process. Following is an analysis of some of these cases and a summary of the principles they establish.

A. *The earliest cases established three broad Constitutional principles.*

First, it was established in *Bayfield County v. Pishon*⁶⁵ that a state could not tax a nonresident on income sourced outside such state.⁶⁶

Second, the U.S. Supreme Court in *Safe Deposit v. Virginia*⁶⁷ held that trust property has "a situs separate and apart

⁶² A partner, S Corporation shareholder, grantor of a trust, or member of an LLC can be taxed on their allocable share of related income even though there has been no distribution and even in the absence of control over such distribution.

⁶³ See *infra* note 227 and accompanying text. In *Lanzi*, [citation], the Division acknowledged the power of ADOR to impose its tax through withholding imposed upon the Alabama Limited Partnership.

⁶⁴ *Hanson v. Denckla*, 357 U.S. 235 (1958); see, e.g., discussion of *Quill*, *infra* note 190.

⁶⁵ *Bayfield County v. Pishon*, 156 N.W. 463 (Wis. 1916).

⁶⁶ The issue was whether Wisconsin could impose a tax on a Wisconsin decedent's testamentary trust's income sourced outside Wisconsin. The trustee, trust property, and beneficiaries were outside Wisconsin. The court treated the trustee as a nonresident. If a resident, source of income would have been irrelevant. The court held on the facts that trust administration within Wisconsin of the Wisconsin settlor's testamentary trust did not result in the trust's income having a Wisconsin source. *Id.*

⁶⁷ *Safe Deposit & Trust Co. of Baltimore, Md. v. Commonwealth of Virginia*, 280 U.S. 83 (1929). The trustee was a resident of Maryland, and the beneficiaries were Virginia residents. Virginia sought to impose an ad valorem tax on trust

from that" of trust beneficiaries.⁶⁸ *Safe Deposit* is cited for the rule that "the taxing power of a state is restricted to her confines and may not be exercised in respect of subjects beyond them."⁶⁹

Third, in *Maguire v. Trefry*⁷⁰, the Supreme Court held that the state of a beneficiary's residence could tax the beneficiary on distributions of current income⁷¹ irrespective of its source.⁷² In *Simmon*,⁷³ the Supreme Judicial Court of Massachusetts ruled, however, that Massachusetts could not tax a Massachusetts resident beneficiary upon previously accumulated trust income when later distributed by a foreign trustee.⁷⁴

In 1935, both federal and Massachusetts statutes only taxed a trust beneficiary when "income,"⁷⁵ as defined by trust accounting rules, was distributed. The definition of "income"

property based upon the beneficiaries' Virginia residency.

⁶⁸ Another early case, *Commissioner of Corporations and Taxation v. Simmons*, 198 N.E. 741 (1935), is discussed *infra* note 114. *Simmons* held a resident of Massachusetts nontaxable on distribution of accumulated income from a nonresident trustee (state income held to have become nontaxable trust principal under Mass. statutes).

⁶⁹ *Guaranty Trust Co. v. Virginia*, 305 U.S. 19, 23 (1938). The Court's reasoning on the risk of double taxation in *Safe Deposit* would probably not be followed today. Fogel, *infra* note 114, at 180-82. *Pennoyer v. Taxation Div. Dir.*, 5 N.J. Tax 386, 395-96 (N.J. Tax Ct. 1983).

⁷⁰ *Maguire v. Trefry*, 253 U.S. 12 (1920). The Massachusetts Supreme Judicial Court upheld and the United States Supreme Court affirmed a Massachusetts tax on the Massachusetts resident beneficiary's income received from the nonresident trust.

⁷¹ *Id.* at 13-14. The case involved current income distributed to the beneficiary, not accumulated income as in *Simmon*, *McCulloch*, and *Gavin*, discussed in notes 73, 23, and 6, respectively.

⁷² The plaintiff beneficiary was a Massachusetts resident. The trust was a testamentary trust of a Pennsylvania resident being administered under Pennsylvania law and the income producing property was in Pennsylvania. *Id.*

⁷³ 198 N.E. 741 (1935). See (first name) Ferrell, *State Taxation of Income Accumulated in Trusts*, 51 A.B.A. J. 566 (June 1965).

⁷⁴ "Foreign" meaning the trust had a nonresident settlor, a nonresident trustee, trust property was outside Massachusetts, and both trust law and trust administration were outside Massachusetts.

⁷⁵ Federal D.N.I rules did not come into existence until the 1940's and 1950's. See Joseph W. Blackburn, *Unique Alabama Trust and Estate Income Tax Rules Create Traps for Alabama Lawyers*, 60 Ala. L. 249 (July, 1999).

included only current year income. Previously accumulated income was treated as having become inextricably commingled with principal.⁷⁶ Thus, the foreign trustee's distribution of accumulated income was treated as a distribution of principal by Massachusetts tax statutes.⁷⁷ A beneficiary's receipt of principle from an estate or trust is viewed as a gift, bequest, or return of capital. Nevertheless, receipt of a gift or bequest could itself constitute "income" for tax purposes.

Income has been defined by the U.S. Supreme Court as an "undeniable accession to wealth clearly realized and over which the taxpayer has complete dominion."⁷⁸ Therefore, there is no Constitutional barrier to taxation of the receipt of property whether state law treats it as "income" or "principal." However, federal and Massachusetts tax statutes exclude gifts and inheritances from taxable income.⁷⁹ Thus, Massachusetts could constitutionally have taxed the Massachusetts beneficiary on the distribution, but statutorily had not taxed it.⁸⁰

The *Simmon* opinion was rendered shortly before Professor Roger John Traynor wrote his classic article on taxation of trusts, settlors, and beneficiaries.⁸¹ Thereafter, as recommended by Professor Traynor, deficiencies in California principal and income statutes were remedied so as to enable California to later tax its resident beneficiaries upon actual distribution of previously accumulated trust income.⁸²

Thus, for the better part of a century it has been clear that a state can tax its *residents* on all income irrespective of source. The issue in *Simmon* was the state's own principal

⁷⁶ See *infra* note 270.

⁷⁷ The opinion cited MASS. GEN. LAWS ANN. Ch. 62, §§ 1, 10 (West 2001).

⁷⁸ *Comm'r v. Glenshaw Glass*, 348 U.S. 426, 75 S. Ct. 473 (1935).

⁷⁹ See, e.g., 26 U.S.C. § 102 (2000).

⁸⁰ A revision of Massachusetts's principal income statutes to redefine income or elimination of the income tax exclusion for gifts and inheritance taxes would have empowered the state to tax its own resident.

⁸¹ See *infra* note 147; Traynor referenced *Simmon*, 198 N.E. 741 (Mass. 1935).

⁸² See *infra* notes 266, 303 and accompanying text. See CAL. REV. & TAX CODE §17745(c) ("... [I]ncome accumulated by a trust continues to be income even though the trust provides that the income (ordinary or capital) shall become a part of the corpus.").

and income statutes which prevented taxation of residents on receipt of accumulated trust income, treating it as a nontaxable principal distribution. Nonresidents can only be taxed on income sourced within the taxing jurisdiction.

B. Later cases more sharply defined the issues.

In *Mercantile-Safe Deposit & Trust Co. v. Murphy*,⁸³ a New York resident created an *inter vivos* trust governed under Maryland law with a Maryland trustee and a New York beneficiary.⁸⁴ After Grantor's death⁸⁵, New York undertook to directly tax the Maryland trustee on accumulated trust income.⁸⁶

The Court of Appeals of New York affirmed the Appellate Division's judgment in favor of the trustee and against levy of the New York tax. The court reasoned that the inability of New York to tax the Maryland trustee was not due to the risk of double taxation. Rather, it was due to "the inability of a State to levy taxes beyond its border."⁸⁷

This case makes clear the rule that a state lacks the power to tax a trustee when the state lacked *in personam* jurisdiction over the trustee and lacked *in rem* jurisdiction over the trust property. There was no other contact with such state and the mere residency of the beneficiary and settlor were not deemed to be contacts of the nonresident trustee.⁸⁸

The court cites *Mercantile Safe Deposit & Trust Co.* for the proposition that:

⁸³ *Mercantile-Safe Deposit & Trust Co. v. Murphy*, 203 N.E.2d 490 (N.Y. 1964).

⁸⁴ *Id.* at 490. *But see infra* note 132 and accompanying text (discussing *District of Columbia v. Chase Manhattan Bank*, 689 A.2d 539 (D.C. 1997)).

⁸⁵ The Grantor's wife, a New York resident, succeeded him as income beneficiary, but income distributions by the Maryland trustee to the wife were merely discretionary. The Trustee was domiciled in Maryland, where the trust, being a non-testamentary, *inter vivos* trust, was administered, and the assets were in the trustee's possession. When the grantor died, his will was probated in New York. *Mercantile-Safe Deposit & Trust Co.*, 203 N.E. 2d. at 490.

⁸⁶ *Id.* at 491.

⁸⁷ *Id.*

⁸⁸ *Id.*

"the imposition of a tax in the State in which the beneficiaries of a trust reside, on securities in the possession of the trustee in another State, to the control or possession of which the beneficiaries have no present right, is in violation of the Fourteenth Amendment."⁸⁹

At this point in the historical review, it is clear that residence of the settlor and/or beneficiary in a state does not cause such state to have either *in rem* or *in personam* jurisdiction over trust property or the trustee. Indeed, residence of the settlor and/or the beneficiary is wholly irrelevant in establishing jurisdiction⁹⁰ as to the nonresident trustee or trust property.⁹¹

In *Pennoyer v. Taxation Division Director*,⁹² a New Jersey domiciliary's will created a testamentary trust.⁹³ The trustee designated by the will was a resident of New York, assets were maintained in New York⁹⁴, and the beneficiaries were not residents of New Jersey.⁹⁵ Taxes were assessed against the trustee by New Jersey. The New Jersey Tax Director claimed the availability of New Jersey courts and the creation of the trust relationship through probate⁹⁶ of decedent's will in New Jersey were sufficient to uphold the tax.⁹⁷

The opinion⁹⁸ discussed several state court cases that had addressed the question.⁹⁹ The court stated that the protection

⁸⁹ *Mercantile-Safe Deposit & Trust Co.*, 203 N.E.2d at 490.

⁹⁰ Reference to *in personam* jurisdiction over the trustee is only as to matters pertaining to the particular trust relationship at issue. See *infra* discussion of *Shaffer* note 231 and accompanying text.

⁹¹ Residence of the settlor is only relevant if, due to default in planning, law of the state of residence governs the trust. *Infra* note 104.

⁹² *Pennoyer v. Taxation Division Director*, 5 N.J. Tax 386 (N.J. Tax Ct. 1983).

⁹³ *Id.* at 390.

⁹⁴ *Id.* at 391.

⁹⁵ *Id.* at 390-91; see also *supra* note 45.

⁹⁶ Probate merely affected the transfer of testator's property to the trustee. Creation and governance of the trust occurs under laws designated in the instrument itself, i.e. New York, not New Jersey. See *supra* note 44.

⁹⁷ *Id.* at 392.

⁹⁸ The court's opinion noted that the United States Supreme Court had not addressed the issue of taxability of a trust on undistributed trust income. *Id.* at 393. But see *supra* note 89.

⁹⁹ *Id.* at 393. The court discusses *Guaranty Trust Co. v. Virginia*, 305 U.S. 19

and control a state exercises over property could give a state the right to tax the income of that property.¹⁰⁰ Ultimately, the court held, "[t]he creation of the trust in 1971 through the probate process in New Jersey courts is an [sic] historical fact which, absent continuing contacts, is not a constitutional nexus justifying income taxation of undistributed income earned in 1979-80."¹⁰¹

Based on *Pennoyer*,¹⁰² a state cannot tax a trustee based solely on historical contacts technically established through transfer of property by a resident to a nonresident.¹⁰³ The trust must actually be created and administered under New Jersey law and must continue to receive significant benefits under New Jersey law. Compare this holding with *District of Columbia v. Chase Manhattan Bank*, which upheld *in personam* jurisdiction over the trustee when the resident decedent's will *did not designate* foreign applicable law or foreign administration. D.C. law and D.C. administration applied to the non-resident trustee's trust relationship in the absence of such a designation.¹⁰⁴

(1938); *Safe-Deposit & Trust Co. v. Virginia*, 280 U.S. 83 (1929); *Bayfield County v. Pishon*, 156 N.W. 463 (Wis. 1916); and *Taylor v. State Tax Comm'n*, 445 N.Y.S.2d 648 (N.Y. App. Div. 1981).

¹⁰⁰ In rem jurisdiction. However, trust property was held in New York by the New York trustees. *Pennoyer*, 5 N.J. Tax at 397.

¹⁰¹ Lack of in personam jurisdiction over the trustee. *Id.* at 397-98. "Since the trust in *not* administered in New Jersey and the trustee, trust assets and beneficiaries are *not* located in New Jersey, the only contact between this trust and New Jersey is the fact that the grantor was domiciled here, letters were issued here, the trustee is amenable to service of process here, and the courts are available to resolve disputes relating to the trust (emphasis added)."

¹⁰² *Supra* note 92.

¹⁰³ Again, a state's statute of wills merely controls effectiveness of the transfer of a decedent/resident's property into the trust. If the will designates a foreign trustee, foreign administration, and foreign applicable law, such foreign law establishes and controls the fiduciary relationship. See *supra* note 44. If New Jersey law provided for a lifetime transfer of a resident's personal property to an out of state transferee, such transferee, whether a trustee or otherwise, does not become subject to *in personam* jurisdiction in New Jersey, nor does New Jersey have *in rem* jurisdiction because of the transfer.

¹⁰⁴ *Infra* note 147 and accompanying text. In *Potter v. Taxation Division Director*, 5 N.J. Tax 399 (N.J. Tax Ct. 1983), a New Jersey resident created an *inter vivos* trust. The trust instrument designated New York law to govern trust ad-

Thus, in *Mercantile*, *Pennoyer*, and *Potter*,¹⁰⁵ state courts in New York and New Jersey had ruled that domicile of a trust's settlor, standing alone, was an insufficient basis for a state to have *in personam* jurisdiction over a nonresident trustee or in *rem* jurisdiction over all trust assets. In *Mercantile*, the added presence of the trust's sole beneficiary failed to establish *in personam* jurisdiction over a nonresident trustee when trust property,¹⁰⁶ applicable trust law, and trust administration were all elsewhere. It was clear from this line of cases that domicile of the settlor and of the beneficiary was not sufficiently related to production of trust income or to contacts of the trustee to permit taxation of the nonresident trustee and property. In *Pennoyer*, it was held that mere probate of a resident's will did not create *in personam* or *in rem* jurisdiction in decedent's state of residence.

In, *Swift v. Director of Revenue*,¹⁰⁷ Missouri courts considered what more was required than settlor/decedent's domicile. Missouri sought to impose a tax on testamentary trusts created by will of a Missouri resident. The trustees, trust property,

ministration. All trust assets and trustees were located outside New Jersey. The trust was designed as a "pour-over" trust, which would receive assets as a beneficiary under the settlor's will. Although New Jersey law applied to effectiveness of the will's transfer of property to the trust, New York law applied to the trust relationship itself. Following the settlor's death, New Jersey claimed it had sufficient contacts to impose a tax on the New York trustee based on the fact that the settlor was domiciled in New Jersey at the time of the trust's creation and at the time of the settlor's death. The court held that the contacts between New Jersey and the *inter vivos* trust were even less than those argued in *Pennoyer*. Indeed, the existing pour-over trust, previously established as an *inter vivos* trust and not as a testamentary trust funded through New Jersey probate, was no different than any other beneficiary under decedent's will. The *Potter* court relied on the reasoning of *Pennoyer* in disallowing the tax. Again, this opinion is distinguished from *District of Columbia v. Chase* (See *infra* note 143) in which resident state law applied to the trust. The distinction between *inter vivos* and testamentary trusts should be irrelevant.

¹⁰⁵ See *supra* notes 82, 92, 104.

¹⁰⁶ Presence of trust property should only be relevant as to in rem jurisdiction over such property and income therefrom. Location of property is irrelevant in establishing *in personam* jurisdiction over its owner, whether the owner is considered to be the trustee or the "trust" as an abstract relationship.

¹⁰⁷ *Swift v. Director of Revenue*, 727 S.W.2d 880 (Mo. 1987) (en banc).

beneficiaries, and the trust administration were all outside Missouri.

The court held that the trusts¹⁰⁸ were not afforded any protection or benefit under Missouri law that constituted sufficient nexus to uphold the tax. The tax levy failed the Due Process test of the Fourteenth Amendment.¹⁰⁹ The *Swift* court's holding follows the holdings in *Mercantile*, *Pennoyer*, *Potter*, and *Taylor*.¹¹⁰ However, the opinion then suggested "relevant" factors to be used to guide the decision-making process in the future. The court considered the following six factors to be relevant in determining whether a sufficient nexus existed to support general *in personam* jurisdiction over the trustee as a Missouri resident¹¹¹:

(1) [T]he domicile of the settlor, (2) the state in which the trust is created, i.e. applicable state law, (3) the location of trust property, (4) the domicile of the beneficiaries, (5) the domicile of the trustees, and (6) the location of the administration of the trust. For purposes of supporting an income tax, the first two of these factors require the ongoing protection or benefit of state law only to the extent that one or more of the other four factors is present.¹¹²

However, it is the thesis of this article that some of these factors are relevant and some are not. Contact number 3, the location of trust property, applies only to the questions of *in rem* jurisdiction on a property by property analysis. Contact number 2 (law governing trust creation and operation), 5 (domicile of trustee), and 6 (location of trust administration) are the only factors relevant to the issue of *in personam* jurisdiction over the trustee. Contacts 1 (settlor domicile) and 2 (bene-

¹⁰⁸ Again, the issue should be whether the nonresident trustee knowingly directed its activities at Missouri, e.g. knowingly became trustee of a trust relationship established and administered under Missouri law.

¹⁰⁹ *Id.*

¹¹⁰ Respectively, *supra* notes 83-106; *infra* notes 212.

¹¹¹ Residency solely as to the trust relationship at issue. Even as a nonresident, the trustee would be taxed on income from property in Missouri as Missouri source income.

¹¹² *Id.* at 882.

ficiary domicile) are only relevant as to a state's taxation of the settlor or the beneficiary.¹¹³

The *Swift* rationale and pronouncement of relevant factors to provide guidelines for the essential "more" for *in personam* jurisdiction has become widely accepted.¹¹⁴ Nevertheless, this article suggests that its listed factors should not have included residence of the settlor or the beneficiary.

The issue in *Westfall v. Director of Revenue*¹¹⁵ was whether a tax may be imposed on all of a resident trust's income when only part of the income was produced from property actually located in the state.¹¹⁶ The settlor was domiciled in Missouri at the time of his death, his will was admitted to probate in Missouri,¹¹⁷ and the trust was subject to Missouri law and administration.¹¹⁸ No income was distributed to Missouri residents, no trustees were residents of Missouri, and no legal proceedings took place in Missouri on behalf of the trustee or the beneficiaries.¹¹⁹

Under the Missouri statute, the entire income of the trustee could be taxed, and credits were available to resident trustees for taxes paid to other states.¹²⁰ No credit was given, how-

¹¹³ See *supra* note 98.

¹¹⁴ Bradley E.S. Fogel, *What Have You Done For Me Lately*, 32 U. RICH. L. REV. 165, 206 (1998); M. Read Moore & Amy L. Silliman, *State Income Taxation of Trust: New Case Creates Uncertainty*, 24 EST. PLN. 200 (June 1997).

¹¹⁵ *Westfall v. Dir. of Revenue*, 804 S.W.2d 27 (Mo. Ct. App. 1990), *aff'd*, 812 S.W.2d 513 (Mo. 1991).

¹¹⁶ *Id.* at 28; see also *Taylor*, *infra* note 211.

¹¹⁷ Wills are always admitted to probate in the state of decedent's domicile. Again, probate merely effects the transfer of property into the trust, but, if the decedent so directs, the trust can be created under the laws of a different state. Probate does not equate to trust creation. A trust is created under applicable laws designated by its Settlor and in the absence of designation, domicile state laws apply.

¹¹⁸ *Id.* Note that the last two factors could have been changed if the trust instrument had designated applicable law of a different state and a different location for trust administration. See *Dist. of Columbia v. Chase Manhattan Bank*, 689 A.2d 539 (D.C. 1997).

¹¹⁹ *Westfall*, 804 S.W.2d at 28-29.

¹²⁰ *Id.* at 29. The tax credit would have avoided double taxation and resulted in proper allocation of the tax burden, as required by the Commerce Clause. See *infra* note 203.

ever, because the trustee did not pay tax to any other state.¹²¹ The court looked to the six factors stated in *Swift* to determine the validity of the tax.¹²² The trustee tried to distinguish *Swift*, but the court, based on Missouri contacts, held the tax valid.¹²³ *Westfall* is consistent with and follows *Mercantile*, *Pennoyer*, *Potter*, *Taylor* and *Swift*.

Again, this article's thesis is in full accord with the result, but not the full rationale, of *Swift*.¹²⁴ Location of property in a state clearly establishes in rem jurisdiction over the income derived from such property. Location of property has little or nothing to do with *in personam* jurisdiction over the trustee, however. As in *Chase*,¹²⁵ the trust relationship, by default, was created under Missouri law and administered in Missouri. Therefore, the trustee knowingly accepted its obligations under and subject to Missouri law.

In *Blue v. Department of Treasury*¹²⁶, a Michigan resident created a revocable *inter vivos* trust.¹²⁷ The trustee was a Florida resident, the trust was formed and administered under Florida law, and the trust assets were located in Florida, except for one piece of real estate, which was located in Michigan.¹²⁸ The Michigan real estate did not produce income. The income beneficiary was a Florida resident. Since the settlor was a resident of Michigan, Michigan's "founder" trustee rules treated the trust as a Michigan resident based solely on domicile of the settlor. Michigan assessed taxes on income from trust property accumulated by the trust between 1982 and 1987.

The trial court held the tax invalid under the Due Process Clause.¹²⁹ On appeal, the Michigan Department of Treasury

¹²¹ *Id.*

¹²² *Id.* at 29-30. See *Swift v. Dir. of Revenue*, 727 S.W.2d 880, 882 (Mo. 1987) (en banc). *Swift* is discussed *supra* notes 96-101 and accompanying text.

¹²³ *Westfall*, 804 S.W.2d at 30.

¹²⁴ *Supra* note 107.

¹²⁵ *Infra* note 148.

¹²⁶ *Blue v. Dep't of Treasury*, 462 N.W.2d 762 (Mich. Ct. App. 1990).

¹²⁷ *Id.* at 763.

¹²⁸ *Id.*

¹²⁹ *Id.*

claimed that sufficient nexus existed to support taxation because the trust received the protection and benefits of Michigan laws.¹³⁰ The Michigan Court of Appeals, in disallowing the tax, stated:

The state cannot create hypothetical legal protections through a *classification scheme* whose validity is constitutionally suspect and attempt to support the constitutionality of the statute by these hypothetical legal protections. We analogize the present case to a hypothetical statute authorizing that any person born in Michigan to resident parents is deemed a resident and taxable as such, no matter where they reside or earn their income. We believe this would be clearly outside of the state's power to impose taxes.

The court followed *Swift and Mercantile-Safe Deposit* in holding the tax invalid under the Due Process Clause.¹³¹

Blue clearly supports the thesis of this article. The Michigan property was clearly subject to the *in rem* jurisdiction of the State of Michigan. The property produced no income, however, and its presence established no basis whatsoever for *in personam* jurisdiction over the Florida trustee. The trust relationship, knowingly accepted by the Florida trustee, was created and administered under Florida law.¹³²

IV. SUMMARY OF PRE-QUILL PRECEDENT

Certain broad pre-Quill principles were clear. States could tax their own residents on their worldwide income.¹³³ Just as

¹³⁰ *Id.* at 764. Note that the Settlor received the benefits of Michigan law, but not the trustee.

¹³¹ *Id.* at 764.

¹³² One author argues that *Blue* and *Westfall* are "inconsistent" because *Blue*, too, technically meets the first three *Swift* factors. The major distinction that can be made of *Blue* is that the Michigan property of the trust was non-income producing. The *Blue* court was fair in its decision and justified in using the strong language quoted, *supra*, to admonish states for finding "hypothetical legal protections" and "constitutionally suspect" classifications for imposing a tax. *Blue*, 462 N.W.2d at 762.

¹³³ See *supra* note 70; see also *Cook v. Tait*, 265 U.S. 47, 55 (1924) (stating, "the native citizen who is taxed may have domicile, and the property from which

clearly, states could not impose taxes, i.e. lacked *in rem* or general *in personam* jurisdiction over income or persons beyond their borders, but could tax income sourced from property within the taxing state. However, as usual, the devil is in the details.

In *Mercantile, Blue, Pennoyer* and *Potter*,¹³⁴ New York, Missouri, and New Jersey all held that mere residency of a settlor, i.e., a Founder Trust, was an insufficient basis for a state to impose general *in personam* jurisdiction over and tax a nonresident, *inter vivos* trustee.¹³⁵ In *Mercantile*, the residency of the income beneficiary in New York, in addition to the Settlor's residency failed to uphold New York's attempt to tax the out-of-state trustee on undistributed trust income held for the benefit of New York residents.¹³⁶

The rulings in *Pennoyer, Mercantile, Swift, Blue, Westfall, Taylor* and *Potter* are well-established and well-reasoned. First, they appropriately reflect the legal distinction between the trustee and the settlor and/or beneficiaries as separate legal persons. Standing alone, residence of the settlor was clearly found to provide insufficient nexus for taxation of a nonresident trustee.¹³⁷

The state of the settlor's domicile could, under various theories, tax the resident settlor on trust income. Grantor trust rules¹³⁸ are a clear example of income tax theories under

his income is derived may have situs, in a foreign country and the tax be legal—the government having power to impose the tax”).

¹³⁴ In *Chase Manhattan Bank v. Gavin*, 733 A.2d 782 (Conn. 1999), the court refused to follow these cases.

¹³⁵ The nexus was insufficient to treat the *inter vivos* trustee as a “resident.” As a nonresident, the trustee could still be taxed on income sourced within the state or as the result of the trustee being subject to general *in personam* jurisdiction.

¹³⁶ Distributed income could, of course, have been taxed by New York to its resident income beneficiary. Even income initially accumulated within the trust could have been taxed by New York to its resident beneficiary when ultimately distributed. See, e.g., *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412, 414-15 (Cal. 1964).

¹³⁷ See Bernard E. Jacob, *An Extended Presence, Interstate Style: First Notes on a Theme from Saenz*, 30 HOFSTRA L. REV. 1133 (2002), for a critical scholarly analysis of “Founder-State Trusts.”

¹³⁸ I.R.C. §§ 671-679 (2000). In particular, see § 679, which applies grantor

which a resident settlor can be taxed on a foreign trust's accumulated income. But residence of the settlor will not provide nexus to allow taxation of a nonresident trustee or claim the trustee is a resident of the taxing state.

Likewise, domicile of a trust's beneficiary clearly establishes a state's nexus to tax the beneficiary but provides no basis whatsoever for taxation of the legally distinct trustee. Beneficiaries can, of course, be taxed upon distributed income or even on accumulated trust income when distributed at a future date.¹³⁹ Taxation of the beneficiary on accumulated trust income before ultimate distribution would be more an issue of realization than of nexus.¹⁴⁰

Mere domicile of the settlor and/or the beneficiary should provide no nexus for treating the trustee as a resident or subject to *in personam* jurisdiction as to the trust relationship. Obviously, the trustee is a legal person separate and distinct from the settlor and/or beneficiary. The trustee must have created nexus through its own contacts directed at the taxing state.¹⁴¹ A resident settlor's establishment of a foreign *inter vivos* trust, *i.e.*, applicable foreign law, foreign administration, and foreign trustee, provides no nexus between the foreign trustee and settlor's state of residence.¹⁴² Establishment of a similar foreign¹⁴³ testamentary trust also provides no nexus with the foreign trustee.

Relevant factors affecting *in personam* jurisdiction over the trustee as to the trust relationship include only: (1) the trust relationship is governed by state law¹⁴⁴; (2) a trustee lives in

trust rules to foreign (non-U.S.) trusts with U.S. beneficiaries.

¹³⁹ See discussion of *Simmon*, *supra* note 73 and accompanying text; see also discussion of *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412 (Cal. 1964), *infra* Part IX and federal throwback rules, discussed *infra* note 265.

¹⁴⁰ See *infra* note 178 and accompanying text.

¹⁴¹ The fiduciary must itself have directed relevant activities at the taxing state as required by *Quill*. See *infra* notes 161, 192.

¹⁴² In *Dist. of Columbia v. Chase Manhattan Bank*, 689 A.2d 539 (D.C. 1997), D.C. law and D.C. administration were, by default, applicable law to the trust, even though trustee, trust assets and beneficiaries were outside of D.C.

¹⁴³ The term "Foreign" meaning applicable foreign law, foreign administration, and a foreign trustee.

¹⁴⁴ In *Swift*, this factor is confusingly described as "the state in which the

the state; or (3) the trust is administered in the state.¹⁴⁵ Situs of property is only relevant as to *in rem* jurisdiction and is not relevant to establish *in personam* jurisdiction over the trustee. Residency of the settlor or beneficiary is not a relevant factor.

In *District of Columbia v. Chase Manhattan Bank*,¹⁴⁶ a post-*Quill* case, a D.C. tax on a foreign trustee was upheld. The trust relationship was established by the will of a resident of D.C., subject to D.C. law, and administered in D.C.¹⁴⁷ The opinion noted that a trust "created" under D.C. law and with D.C. administration, much like a domestic corporation, was a creature of the law of its state of organization.¹⁴⁸

trust is created." This terminology confuses the issue of probate and compliance with a state's statute of wills [wholly irrelevant factors] on one hand, and the law which governs establishment and operation of the trust relationship on the other.

¹⁴⁵ IDAHO ADMIN. CODE 35.1.1.35 (2003). See *Income Taxation of Trust Income: Idaho Law, Practical Drafting* 6618-6619 (October, 2001).

¹⁴⁶ *Dist. of Columbia v. Chase Manhattan Bank*, 689 A.2d 539 (D.C. 1997).

¹⁴⁷ Professor Roger John Traynor had predicted taxation of testamentary trusts on this basis many years ago in his classic article on trust taxation. See Roger John Traynor, *State Taxation of Trust Income*, 22 IOWA L. REV. 268 (1937) (discussed *infra* Part VIII).

¹⁴⁸ *Chase Manhattan Bank*, 689 A.2d at 544. The will did not designate foreign applicable law or foreign administration, so, by default, D.C. law applied. See *supra* note 44. The opinion refused to extend its holding to an *inter vivos* trust established by a D.C. resident as follows:

We express no opinion as to the constitutionality of taxing the entire net income of *inter vivos* trusts based solely on the fact that the settlor was domiciled in the District when she died and the trust therefore became irrevocable. In such cases, the nexus between the trust and the District is arguably more attenuated, since the trust was not created by probate of the decedent's will in the District's courts. An irrevocable *inter vivos* trust does not owe its existence to the laws and courts of the District in the same way that the testamentary trust at issue in the present case does, and thus it does not have the same permanent tie to the District. In some cases the District courts may not even have principle supervisory authority over such an *inter vivos* trust. The idea of fundamental fairness, which under girds our due process analysis, therefore may or may not compel a different result in an *inter vivos* trust context. (Emphasis added). Everything the court stated as to an *inter vivos* trust is equally applicable to a testamentary trust when the will designates foreign law and foreign administration.

Id. at 547 n.11. Again, in the "present case" the will failed to invoke foreign law or foreign administration.

The *Chase* opinion's finding of sufficient nexus was based on the specific facts of that case. Despite dictum¹⁴⁹, nexus was not based on mere residency of the decedent in the District. Neither was nexus founded on mere probate of decedent's will in the District. Nexus was based on the trust relationship being administered in the District under District laws. Decedent had not designated administration or laws of another jurisdiction. The foreign trustee accepted its duties subject to such conditions.

Other than the unprecedented holding in *Gavin*, none of these cases found due process nexus based solely on the residence of the settlor and/or beneficiary in the taxing state.

V. *QUILL CORP. V. NORTH DAKOTA*¹⁵⁰

Quill Corporation conducted a mail order business throughout the United States. Although Quill was not a resident of North Dakota and had no physical presence in the state, *i.e.*, no warehouse, office, etc., it did make catalogue sales of products into North Dakota.¹⁵¹ The United States Supreme Court addressed the validity of the North Dakota use tax on a nonresident under both the Due Process Clause and the Commerce Clause of the U. S. Constitution.¹⁵²

In addressing North Dakota's ability to tax Quill's North Dakota activities¹⁵³ within due process limits, the Court noted that the Due Process Clause required: (1) "some definite link, some minimum connection, between a state and a person, property or transaction it seeks to tax,"¹⁵⁴ and (2) "that the income attributed to the State for tax purposes must be rationally related to values connected with the taxing State."¹⁵⁵ The

¹⁴⁹ *Id.*

¹⁵⁰ 504 U.S. 298.

¹⁵¹ *Id.* at 302.

¹⁵² *Id.* at 305.

¹⁵³ North Dakota never attempted to tax Quill as a "resident."

¹⁵⁴ *Id.* (quoting *Miller Bros. Co. v. Md.*, 347 U.S. 340, 344-45 (1954)).

¹⁵⁵ *Id.* (quoting *Moorman Mfg. Co. v. Blair*, 437 U.S. 267 (1978)). Quill was clearly being taxed as a nonresident. The North Dakota tax would only be imposed on Quill's North Dakota sales. Quill would not be subject as a resident

North Dakota tax being levied in this case was a use tax. Relying on *International Shoe Company v. Washington*¹⁵⁶ and its progeny, the Court stated, "we have framed the relevant inquiry as whether a defendant had minimum contacts with the jurisdiction 'such that the maintenance of the suit does not offend traditional notions of fair play and substantial justice.'"¹⁵⁷

The Court continued its analysis and held that due process could be satisfied even if a taxpayer had no physical presence within the taxing state.¹⁵⁸ Regarding the imposition of a sales/use tax, the Court held that such reasoning "justifies the imposition of the collection duty on a mail order house that is engaged in continuous and widespread solicitation of business within a state. Such a corporation clearly has 'fair warning that [its] activity may subject [it] to the jurisdiction of a foreign sovereign.'"¹⁵⁹ Thus, North Dakota's sales/use tax levy was held to satisfy the Court's due process standard.

Analysis of the negative sweep of the Commerce Clause, *i.e.*, those aspects of state activity prohibited because they may unduly interfere with interstate commerce, ultimately led to a determination that the sales/use tax could not be imposed by North Dakota. The Court analyzed the negative impact of state taxation on interstate commerce. The Court then noted that it had "adopted a 'multiple-taxation doctrine' that focused not on whether a tax was 'direct' or 'indirect,' but rather on whether a tax subjected interstate commerce to "*a risk of multiple taxation*."¹⁶⁰

The Court ultimately held that it would apply the four-part test which it first established in *Complete Auto Transit, Inc. v. Brady* to the Commerce Clause question.¹⁶¹ The Court stated:

taxpayer to general taxation by North Dakota on all its sales. *Id.* at 303.

¹⁵⁶ *Int'l Shoe Co. v. Wash.*, 326 U.S. 310 (1945).

¹⁵⁷ *Quill*, 504 U.S. at 307 (citations omitted).

¹⁵⁸ *Id.* at 308.

¹⁵⁹ *Id.* (quoting *Shaffer v. Heitner*, 433 U.S. 186, 218 (1977) (citations omitted)).

¹⁶⁰ *Id.* at 309 (emphasis added).

¹⁶¹ *Quill*, 504 U.S. at 310 (citing *Complete Auto Transit, Inc. v. Brady*, 430 U.S. 274, 279 (1977)).

Under *Complete Auto's* four-part test we will sustain a tax against a Commerce Clause challenge so long as the "tax [1] is applied to an activity with a *substantial nexus* with the taxing State, [2] is fairly apportioned, [3] does not discriminate against interstate commerce, and [4] is fairly related to the services provided by the State."¹⁶²

Again, this test was being applied to the allocable, i.e. North Dakota source, sales of a nonresident.¹⁶³ Neither *Complete Auto* nor *Quill* dealt with classification and taxation of a resident on its worldwide income. Even if state law denominates or classifies a trust as being a "resident trust," imposition of a tax on the trustee must not violate due process or the Commerce Clause of the U.S. Constitution.¹⁶⁴

At least in the context of sales/use taxes and the mail order business, the Court reaffirmed the continued applicability of the bright line presence test set forth in *National Bellas Hess, Inc. v. Department of Revenue of Ill.*¹⁶⁵ as applied to factor [1] of *Complete Auto's* four-part Commerce Clause test.¹⁶⁶ The Court was also emphatic that the due process "minimum contacts" test required much less than the "substantial nexus" test of the Commerce Clause, stating, "[a] tax may be consistent with due process and yet unduly burden interstate commerce."¹⁶⁷ The Court held that the first and fourth prongs of the *Complete Auto* test "require a substantial nexus and a relationship between the tax and state-provided services,"¹⁶⁸ to

¹⁶² *Id.* at 311 (citation omitted) (emphasis added).

¹⁶³ *Id.* at 298.

¹⁶⁴ *Id.* at 305. See *supra* note 135 and *Blue*, 462 N.W.2d 762, regarding "constitutionally suspect classifications."

¹⁶⁵ *Nat'l Bellas Hess, Inc. v. Dep't of Revenue of Ill.*, 386 U.S. 753, 765 (1967).

¹⁶⁶ *Quill*, 504 U.S. at 317 ("[T]he continuing value of a bright-line rule in this area and the doctrine and principles of *stare decisis* indicate that the *Bellas Hess* rule remains good law."). The *Bellas Hess* test remains good, at least as applied to state sales and use taxes and the mail order business, but the Supreme Court also stated that "the bright-line rule of *Bellas Hess* furthers the ends of the dormant Commerce Clause." *Id.* at 314.

¹⁶⁷ *Id.* at 313 n.7.

¹⁶⁸ *Id.* at 313.

ensure that state taxation does not unduly burden interstate commerce.¹⁶⁹

VI. ARE STATE BENEFITS AFFORDED A BENEFICIARY ATTRIBUTABLE TO A NONRESIDENT TRUSTEE FOR DUE PROCESS PURPOSES?

A. Gavin

The facts of *Gavin* were summarized in the Introduction to this article. Connecticut taxed a New York trustee bank based on claimed in personam jurisdiction as to the trust relationship.¹⁷⁰ The trustee had no Connecticut contacts, trust property was in New York, trust administration was in New York,¹⁷¹ and New York law governed the creation and enforcement of the trust relationship.¹⁷²

The *Gavin* opinion justified its taxation of the foreign trustee of the single inter vivos trust by citing benefits Connecticut law provided to the trust's Connecticut beneficiary.¹⁷³ Based on this analysis, Connecticut would tax, and claim compliance with due process, any estate or trust, whether testamentary or inter vivos,¹⁷⁴ established by even a nonresident settlor, if such estate or trust had a Connecticut beneficiary.¹⁷⁵ The following benefits to the *Gavin* beneficiary cited by

¹⁶⁹ *Id.* The "[p]urpose of the Commerce Clause is to insure a national economy free from such unjustifiable local entanglements. . . ." *Bellas Hess*, 386 U.S. at 760. Realize that the first and fourth prongs of this test address the same two issues as due process analysis, i.e. (1) sufficiency of nexus, and (2) relationship between state tax and state benefits provided. See *supra* note 156.

¹⁷⁰ *Chase Manhattan Bank v. Gavin*, 733 A.2d 782, 790 (Conn. 1999).

¹⁷¹ *Id.* at 803-04.

¹⁷² *Id.* at 786.

¹⁷³ *Id.* at 803-04. Minimum contacts were deemed satisfied by "the presence in the state of a noncontingent beneficiary who receives the benefits and protections of the state's laws." *Id.* at 803.

¹⁷⁴ *Id.* at 801. Regarding inter vivos trusts, the *Gavin* court stated "[a]lthough this is a closer case than that with respect to the testamentary trusts, we conclude that the state's taxation scheme of the undistributed income of the inter vivos trust satisfies the Due Process clause." *Id.* at 801.

¹⁷⁵ *Id.* at 802. Although the settlor in *Gavin* was a Connecticut resident when the inter vivos trust was established, that was not, and could not have been, the basis for finding minimum contacts for due process purposes. *Id.* at 802.

the opinion would apply equally to a Connecticut beneficiary of any nonresident estate or trust, whether or not established by a Connecticut settlor:

In the present case, the *critical link* to the undistributed income sought to be taxed is the fact that *the non-contingent beneficiary* of the *inter vivos* trust during the tax year in question *was a Connecticut domiciliary*. The accumulated income eventually will be paid . . . to her . . . [or] to her then living descendants. Thus . . . *she enjoyed* all of the protections and benefits afforded to other domiciliaries. *Her right* to the eventual receipt and enjoyment of the accumulated income was . . . protected by the laws of the state. We conclude that . . . a state may . . . tax the income of an *inter vivos* trust that is *accumulated for the ultimate benefit of a non-contingent domiciliary* ¹⁷⁶

Connecticut law benefited its resident beneficiary.¹⁷⁷ If the beneficiary's noncontingent status satisfied the "realization" requirement for "income," Connecticut could clearly tax its beneficiary.¹⁷⁸ The *Gavin* opinion, however, improperly attributed such benefits to the New York trustee for purposes of due process. Again, the New York trustee had no contacts with Connecticut and derived no benefits from Connecticut law.¹⁷⁹ Once this attribution from beneficiary to trustee is allowed to occur, there is the potential for endless reattribution of state law benefits, *i.e.*, why stop with mere single attribution? This rationale would and did turn due process on its head in the context of *Gavin*: "Where there is jurisdiction neither as to person nor property, the imposition of a tax would be *ultra vires* and void."

¹⁸⁰ Could shareholder domicile states tax nonresident corporations that have no business or other contact with such

¹⁷⁶ *Id.* at 802 (emphasis added). See discussion *infra* note 228.

¹⁷⁷ *Id.* at 802.

¹⁷⁸ See discussion of *Simmon* and *Glenshaw Glass*, *supra* note 73.

¹⁷⁹ See *supra* notes 9-13 and accompanying text.

¹⁸⁰ *Miller Bros. Co. v. Maryland*, 347 U.S. 340, 342 (citing *City of St. Louis v. Wiggins Ferry Co.*, 78 U.S. (11 Wall.) 423, 430 (1870)).

state, or vice versa, as a result of attribution of state law benefits from a resident shareholder to such corporation? We know the answer to this rhetorical question—it is clearly No!¹⁸¹

In *Cannon Manufacturing Co. v. Cudahy Packing Co.*,¹⁸² the Supreme Court first held that corporate contacts cannot be attributed to shareholders of the corporation and vice versa.¹⁸³ This principle has clearly been extended to a limited partnership and attribution of its contacts to limited partners.¹⁸⁴ If general partners are agents acting on behalf of their partnership, contacts of the general partner can be attributed to the partnership.¹⁸⁵ This is solely because the partner's actions as agent are actions of the partnership.¹⁸⁶ However, beneficiaries, as such, are not agents of the trustee. Clearly, the beneficiary and the trustee are distinct legal persons.¹⁸⁷

In *Quill*, the Supreme Court looked to the taxpayer's direct contacts with and benefits derived from North Dakota.¹⁸⁸ Indirect benefits of North Dakota law to Quill's North Dakota customers or vendors which might be attributed to Quill were not even analyzed by the Court.¹⁸⁹ To establish "minimum contacts" for due process, the Supreme Court noted that Quill had mailed twenty-four tons of sales catalogs into the state and had made sales there of approximately one million dollars.¹⁹⁰ As a result, Quill was found to have "purposefully directed its

¹⁸¹ See *infra* note 184 and accompanying text.

¹⁸² 267 U.S. 333 (1925).

¹⁸³ *Id.* at 336. A subsidiary corporation's operations in North Carolina failed to establish due process nexus between North Carolina and the subsidiaries' parent corporation, i.e. its shareholder. Separateness of the subsidiary was not a "pure fiction" and was respected. *Id.* at 337.

¹⁸⁴ *Sher v. Johnson*, 911 F.2d 1357, 1366 (9th Cir. 1990).

¹⁸⁵ *Id.*

¹⁸⁶ *Id.* For a well-reasoned analysis of the fallacies of nexus attribution, see Lowery & Vasquez, "When Is It Unconstitutional For States To Tax Nonresident Members of Limited Liability Companies?" *State Tax Notes*, May 19, 2003, at 633.

¹⁸⁷ See *supra* notes 28-35 and accompanying text.

¹⁸⁸ *Quill Corp. v. North Dakota*, 504 U.S. 298, 308 (1992).

¹⁸⁹ See *infra* note 194.

¹⁹⁰ *Quill*, 504 U.S. at 305.

activities at North Dakota..."¹⁹¹ Quill was not and could not be subjected to tax merely because North Dakota law applied to and benefited Quill's customers or vendors and thereby indirectly benefited Quill.¹⁹² Such state benefits were not and could not be attributed to Quill.¹⁹³

B. Gavin's Treatment of Trust Relationships as Forever Establishing Tax Jurisdiction Over the Trustee When Trustees, Trust Property, Trust Administration, and Applicable Trust Law Were all Outside of Connecticut.

State taxation of a trustee, whether the trust relationship was testamentary or inter vivos, must pass both the Due Process and Commerce Clause requirements established under *Quill*. Even if the settlor resided in the taxing state, tax policies which create a risk of multiple taxation nevertheless discriminate against interstate commerce and are unconstitutional.¹⁹⁴

The *Gavin* opinion clearly recognized that taxation of a trustee when the trust relationship was actually established, governed, and administered under the taxing state's applicable law was easily justified.¹⁹⁵

Taxation of the trustee of an inter vivos trust relationship established, governed, and administered outside the taxing state under laws of a foreign jurisdiction can hardly be justified on the same principles, however.¹⁹⁶ Likewise, testamentary trust relationships, like inter vivos trust relationships, can clearly be established, administered, and subject to laws of a

¹⁹¹ *Quill*, 504 U.S. at 308.

¹⁹² *Id.* at 313.

¹⁹³ *Id.* *Quill* focused on the taxpayer's direct contacts with its North Dakota customers, not the mere fact that taxpayer had customers in the state. *Id.* at 313 n.6 (citing *Nat'l Bellas Hess, Inc. v. Dept. of Revenue of Ill.*, 386 U.S. 753, 759-60 (1967)). Indeed, the North Dakota Supreme Court relied on similar, *Gavin*-like, general economic benefits and economic climate provided by its laws, *id.* at 304, but was reversed by the Supreme Court.

¹⁹⁴ See Part B.3.

¹⁹⁵ *Gavin*, 733 A.2d at 800 (citing *Dist. of Columbia v. Chase Manhattan Bank*, 689 A.2d 539, 540 (D.C. Cir. 1997)).

¹⁹⁶ See *supra* note 149.

foreign jurisdiction.¹⁹⁷ Such testamentary trust relationships have no more contact with the jurisdiction of the Decedent's domicile than a similar inter vivos trust relationship.¹⁹⁸

VII. GAVIN'S FAILED COMMERCE CLAUSE ANALYSIS

As stated in *Quill*¹⁹⁹ and *Complete Auto*,²⁰⁰ a state tax will be sustained under a Commerce Clause challenge²⁰¹ only so long as the "tax (1) is applied to an activity with a *substantial nexus* with the taxing State, (2) is fairly apportioned, (3) does not discriminate against interstate commerce, and (4) is fairly related to services provided by the State."²⁰²

1. *Substantial Nexus*. Substantial nexus requires more than "minimum contacts," "slightest presence," or "minimal nexus" under due process.²⁰³ In *Gavin*, out-of-state banks, as trustees, had to have "a substantial nexus" with Connecticut in order for Connecticut to levy an income tax on such out-of-state trustees²⁰⁴. Yet, except in its due process analysis, *Gavin* makes no analysis whatsoever as to sufficiency of the out-of-state bank's contacts as a trustee with the State of Connecticut. As previously discussed,²⁰⁵ the trustee's contacts with Connecticut even failed the minimal due process nexus requirement.

¹⁹⁷ By default, i.e. by failure of the instrument to designate foreign law and foreign administration, District laws applied to the creation and administration of the trust in *Chase*, 689 A.2d at 543. See *supra* notes 147, 149.

¹⁹⁸ See *supra* notes 40-46 and discussion of *Shaffer*, *infra* note 228.

¹⁹⁹ See *supra* notes 150-52 and accompanying text.

²⁰⁰ See *supra* note 161 and accompanying text.

²⁰¹ The Commerce Clause's negative sweep is the source of limitations on a state's power to tax interstate transactions. See *Quill*, 504 U.S. at 309 ("[the Court's] interpretation of the 'negative' or 'dormant' Commerce Clause has evolved substantially over the years, particularly as that Clause concerns limitations on state taxation powers").

²⁰² *Id.* at 311 (citing *Complete Auto Transit, Inc.*, 430 U.S. at 279) (emphasis added).

²⁰³ *Id.* at 315 n.8.

²⁰⁴ Only legal persons can have nexus, i.e. can direct their activities at the taxing state. Trusts are not legal persons. Trusts are merely abstract relationships between persons. See *supra* notes 30-35 and accompanying text.

²⁰⁵ See *supra* notes 192-94.

The *Gavin* opinion merely suggests that the out-of-state bank fails to address the four separate aspects of the Commerce Clause test.²⁰⁶ The opinion itself merely addressed whether risk of multiple taxation would unconstitutionally impact selection of out-of-state trustees as contrasted with domestic trustees not subject to such risk.²⁰⁷

2. *Fair Apportionment.* Fair apportionment of the tax among states involved in the taxed activity is also required.²⁰⁸ As stated by the dissent in *Gavin*, Connecticut's income tax is not apportioned at all.²⁰⁹ There was no tax credit available from Connecticut for the trustees' taxes paid to New York.²¹⁰ Apportionment requires a rational and measured fragmenting of the transaction or tax base so that each state taxes only that part of the whole activity or base most closely related to such state.²¹¹ *Gavin* taxed all of the trustee's income, not merely some apportioned part.²¹²

Trust assets, administration, applicable laws, and trustees were all outside of Connecticut. Nevertheless, the trustee was treated as a "resident" in *Gavin*, i.e. taxed on all their income,

²⁰⁶ *Gavin*, 733 A.2d at 805.

²⁰⁷ See *infra* note 212.

²⁰⁸ *Quill*, 504 U.S. at 311 (citing *Complete Auto Transit*, 430 U.S. at 279).

²⁰⁹ *Gavin*, 733 A.2d at 808. The trusts were all treated as "resident" trusts under *Gavin*, and residents are taxed on all income without allocation. Connecticut did not even allow tax credits. *Id.* at 786-87.

²¹⁰ *Id.* at 804. The trustee and trust property were in New York and New York law applied to trust administration. The *inter vivos* trust was formed under New York law. Also, see *supra* notes 116-24 and accompanying text.

²¹¹ In *Taylor v. State Tax Commission*, 445 N.Y.S.2d 648 (N.Y. App. Div. 1981), a testamentary trust was created by the will of a decedent who died while residing in New York. The trust assets included real property in Florida. Since Florida law prohibited a foreign corporation from acting as trustee over Florida property, decedent's sons were appointed as trustees of the Florida property. The trust excluded capital gains from the Florida property when calculating its New York taxable income. The court held that neither New York laws nor New York services were of benefit to the Florida property. The court stated, "[t]he fact that the former owner of the property in question died while being domiciled in New York, making the trust a resident trust under New York tax law, is insufficient to establish a basis for jurisdiction." Thus, New York taxation of gain on the Florida property violated the Fourteenth Amendment to the U.S. Constitution.

²¹² See *supra* note 15.

irrespective of the source of such income.²¹³

Rational allocation has been attempted by many states in many ways. The "UDITPA" applies uniform allocation rules to all signatory states.²¹⁴ Under the UDITPA there are different allocations for business and for non-business²¹⁵ income. Sales of a business's assets are allocated in one fashion and sale of its stock is allocated in another.²¹⁶ Also, states such as New York have their own unique allocation rules.²¹⁷ Pass-through entities add another layer of complexity to the allocation confusion.

Overly broad²¹⁸ trustee nexus and residency statutes are sometimes referred to conservatively as Founder State Trusts. Even if a trustee were a resident, the right of other states to tax income from sources within such other states must be recognized and adjustments made to avoid the risk of multiple tax burdens.²¹⁹

3. *Discrimination Against Interstate Commerce.* The principle focus of the *Gavin* opinion's Commerce Clause analysis was whether or not a Connecticut "settlor would choose a Connecticut bank, over an out-of-state bank, as trustee *solely* because of the potential for future multiple taxation of some portion of the trust's future income."²²⁰ The appropriate test, however, was whether or not the "practical effect" of the Connecticut tax was to "discriminate against interstate commerce."²²¹

As noted, *supra*, the *Gavin* opinion stated, "Although we agree that there are such incentives and risks, we conclude

²¹³ See *supra* note 202 and accompanying text.

²¹⁴ Unif. Div. of Income for Tax Purposes Act.

²¹⁵ See, e.g., *Ex parte Uniroyal Tire Co.*, 779 So. 2d 227 (Ala. 2000); *Allied-Signal, Inc. v. Dir.*, 504 U.S. 768 (1992).

²¹⁶ *Mobile Oil v. Comm'r of Taxes*, 445 U.S. 425 (1980).

²¹⁷ See N.Y. Tax Law §§ 210.3, 208.9.

²¹⁸ "The Founder-State Trust is inconsistent with the relations of states of the American Union and with the reality, and realistic expectations, of the citizens of each state." Bernard E. Jacob, *An Extended Presence, Interstate Style: First Notes on a Theme from Saenz*, 30 HOFSTRA L. REV. 1133, 1240 (2002).

²¹⁹ See *infra* note 222.

²²⁰ *Gavin*, 733 A.2d at 806 (emphasis added).

²²¹ *Quill*, 504 U.S. at 310-11.

that they are too remote and speculative to constitute a dormant commerce clause violation."²²² The opinion found the speculative nature in the selection of a trustee from its belief that multiple and diverse factors²²³ would be considered by a Connecticut settlor before selecting a trustee. The Connecticut court found that its tax imposed on a foreign trustee would indeed subject the trust income to the risk of multiple taxation,²²⁴ a risk not shared by the domestic trustees. Thus, even if a state tax created an undue burden on interstate commerce, the *Gavin* opinion would not find a Commerce Clause violation unless it could determine the precise role such tax played in commercial decisions of consumers of interstate products and services. Indeed, the opinion stated that the adverse impact of the multiple taxes must be the "sole"²²⁵ determinative factor on economic decisions affecting interstate commerce. In summary, as well-stated by one commentator, "[I]f a state elects to attempt to tax income from trusts having contacts on which the Connecticut Supreme Court supported its tax . . . there are strong arguments to support a challenge to the enforceability of such tax."²²⁶

4. *Relation to State Benefits Provided.* *Shaffer v. Heitner*²²⁷ was decided by the Supreme Court in the same time frame as it decided *Complete Auto*.²²⁸ In *Shaffer*, the Su-

²²² *Gavin*, 733 A.2d at 805. See *infra* note 224 and accompanying text.

²²³ "The choice of bank as trustee is likely to be animated by many imponderables, among them: the prior experience of the testator or settlor; the financial performance of the various banks in the pool of available choices; the current location of the likely beneficiaries; the current location of the various banks in the pool; and, of course, the tax implications, if any. We simply are not sufficiently persuaded of the underlying validity of the plaintiff's assumption about how such a multifaceted decision as the choice of a trustee is likely to be made." *Id.* at 806.

²²⁴ "[B]ecause the property of all of these trusts consisted only of cash and securities in the tax year in question, no credit would have been available to them under the current tax scheme." *Id.* at 804.

²²⁵ See *supra* note 221.

²²⁶ Max Guitierrez, Jr., *The State Income Taxation of Multi-Jurisdictional Trusts*, 36 U. MIAMI INST. ON EST. PLAN. §1309.2 (2002).

²²⁷ 433 U.S. 186 (1977).

²²⁸ See *supra* note 161.

preme Court analyzed personal jurisdiction under the *International Shoe*²²⁹ due process standard.²³⁰ As part of the application of this standard, the Supreme Court also analyzed whether the minimum contacts of a party were related to the subject matter at issue.²³¹ In essence, this became a second due process requirement. "[T]here is no necessary relationship between holding a position as a corporate fiduciary [subject of the litigation] and owning stock or other interest in the corporation [parties' sole due process contact with Delaware]."²³² The Supreme Court also directly rejected a state's claim that benefits provided through application of state laws to an issue, necessarily gave the state jurisdiction over the parties associated with such issue.²³³ "The issue is personal jurisdiction, not choice of law. It is resolved in this case by considering the acts of the [appellants]."²³⁴

(1) Minimum contacts, and (2) relation of such contact to the subject matter at issue, i.e. production of income or subject matter of litigation. In *Heitner*, parties had property in Delaware [thereby satisfying (1) above], but the property was not the subject matter of the litigation in which personal jurisdiction was sought [thereby, not satisfying (2) above].

The Appellee's discredited argument in *Shaffer* was much the same as Connecticut's in *Gavin*: "[Appellee] notes that Delaware law provides substantial benefits to corporate officers and directors, and that . . . [it is] only fair and just to require appellants, in return for these benefits, to respond in the State of

²²⁹ See *supra* note 159.

²³⁰ "Minimum contacts" such as do not offend traditional notions of "fair play and substantial justice." See *supra* note 158; *Shaffer*, 433 U.S. at 203, 207.

²³¹ *Shaffer*, 433 U.S. at 213.

²³² *Id.* at 214.

²³³ *Id.* at 215 ("[W]e have rejected the argument that if a State's law can properly be applied to a dispute, its courts necessarily have jurisdiction over the parties to that dispute.").

²³⁴ *Id.* (citations omitted) (emphasis added). *Quill*, likewise, ignored general benefits derived from applicability of North Dakota law but focused instead on *Quill's* own acts directed at North Dakota. See *supra* note 192.

Delaware."²³⁵ But the Court absolutely rejected this general "legal benefits" argument and stated "this line of reasoning . . . does *not* demonstrate that appellants have 'purposefully availed themselves of the privilege of conducting activities within the forum state.' *Hanson v. Denckla*, 357 U.S. 235, at 253 (1958)."²³⁶

In *Complete Auto's* four-part test for state taxation's compliance with the Commerce clause, Steps One²³⁷ and Four essentially repeat this dual due process standard. Step One requires adequate nexus.²³⁸ Step Four requires the necessary relationship between state benefits provided and the subject of taxation.²³⁹

Gavin wrongly held that the Connecticut law benefits afforded the Connecticut beneficiary²⁴⁰ subjected the nonresident trustee to Connecticut jurisdiction for due process purposes. This analysis was squarely rejected by the Supreme Court in *Shaffer*,²⁴¹ whether the relationship to state services provided test is applied in the due process context or as step four in the Commerce Clause analysis.²⁴²

VIII. THE IMPORTANCE OF ROGER JOHN TRAYNOR²⁴³ TO THE ANALYSIS OF BOTH *MCCULLOCH* AND *GAVIN*

Justice Roger John Traynor was among the most influen-

²³⁵ *Id.* at 215-16 (internal quotations omitted).

²³⁶ *Id.* at 216 (emphasis added).

²³⁷ See *supra* note 168.

²³⁸ See *supra* note 204.

²³⁹ See *supra* note 156.

²⁴⁰ See *supra* note 176. Connecticut law did not benefit the earning of the income from trust property. Connecticut law merely protected the beneficiary's receipt and enjoyment of income when distributed.

²⁴¹ See *Northwestern States Portland Cement Co. v. Minnesota*, 358 U.S. 450, 464-65 (1959), (taxes are "levied only on that portion of the taxpayer's net income which arises from its activities within the taxing State").

²⁴² Seth Goldstein, "Resident" Taxpayers: *Internal Consistency, Due Process, and State Income Taxation*, 91 COLUM. L. REV. 119, 141 (1991).

²⁴³ Professor of Law, University of California, Advisor to the California State Board of Equalization. He "played an important role in drafting important [California] tax laws." National Tax Association Tax Conference, Indianapolis, Indiana, September 29, 1936, at 157.

tial and highly esteemed jurists and legal scholars of the twentieth century.²⁴⁴ From 1928 to 1940, Professor Traynor served as tax advisor to the California Board of Equalization where he was responsible for complete reform of California's sales and use tax, personal income tax, and corporate franchise taxes. He took a leave of absence in 1937 to serve at the U.S. Treasury Department to help draft the Revenue Act of 1938.²⁴⁵ He served as associate justice of the California Supreme Court from 1940 to 1964, serving as chief justice from 1964 to 1970.²⁴⁶

Among his prolific writings, Professor Traynor authored the classic insightful article, *State Taxation of Trust Income*.²⁴⁷ As associate justice of the California Supreme Court, he was in the majority of the Supreme Court of California on its *McCulloch* opinion. Indeed, the special California statutes²⁴⁸ analyzed in *McCulloch* reflected in detail the views expressed by Professor Traynor in his early article dealing with state taxation of trustees, beneficiaries, and settlor/grantors.²⁴⁹

²⁴⁴ In his writing and jurisprudence, he established national legal principles of product liability with strict liability; extension of strict product liability through the distribution chain; action for intentional infliction of emotional distress; "moderate and restrained" doctrine in conflict of laws; the fiduciary obligation owed by majority shareholders to minority shareholders; admission of custom and practice evidence in contract litigation; the "exclusionary rule" barring improperly obtained evidence; abolition of sovereign immunity; no fault divorce through elimination of defense of recrimination; abolishment of laws prohibiting miscegenation; and citation of law review articles in judicial decisions. See generally "Roger John Traynor", Legal Encyclopedia, published by Thomson & Gale; West's Encyclopedia of American Law, The Gale Group, Inc. 1998; Answers.com, Roger J. Traynor, <http://www.answers.com/topic/traynor-roger-john>; see also Henry J. Friendly, *Ablest Judge of His Generation*, 71 CAL. L. REV. 1039 (1983); G. Edward White, *Introduction to The Traynor Reader: A Collection of Essays by the Honorable Roger J. Traynor*, <http://uchastings.edu/?pid=1408>.

²⁴⁵ See *supra* note 246.

²⁴⁶ See *supra* note 246.

²⁴⁷ Roger John Traynor, *State Taxation of Trust Income*, 22 IOWA L. REV. 268 (1937).

²⁴⁸ See *supra* notes 203-07 and accompanying text.

²⁴⁹ Indeed, *McCulloch* cites Professor Traynor's article in footnote nine. See *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412, 420 n.9 (Cal. 1964) ("[W]hile the complexity of the trust itself and of the relations of the parties thereto compli-

Traynor's article first analyzed settled law on state taxation of residents and nonresidents.²⁵⁰ He discussed a state's right to tax domestic trustees on all their accumulated income whether earned within or without its borders.²⁵¹ He also acknowledged the right of a state to tax all income earned within its borders by nonresident trustees, thereby resulting in two states possibly taxing the same income.²⁵² Likewise, he discussed the well-established right of a state to tax resident beneficiaries upon distributions of current income.

Traynor's article, however, then focused on how the complexity of trusts had created obstacles to a state's ability to tax accumulated trust income.²⁵³ Nonresident, complex trusts located in states selected for minimum taxation could and were being used to shelter income.²⁵⁴ Through accumulation of income by a nonresident trustee, the state of the beneficiary's residence could be prevented from taxing such income.²⁵⁵ *Simmon*²⁵⁶ clearly reflected how a subsequent distribution of previously accumulated trust income could also avoid taxation by the beneficiary's state of residence depending on a state's

cates the problem of effective taxation, it should not obstruct the claims of a state to tax trust income, so far as possible, as it income is accorded of a trust. Trust income is accorded protection in its production, receipt and enjoyment to the same extent as other income; it measures in like manner ability to pay. It measures in like manner ability far as possible upon the same basis. If the obstacles interposed by the trust device are to be circumvented, jurisdiction to tax should be found wherever substantial claims to tax are reinforced by effective power to compel payment." See Traynor, *supra* note 247, at 271- 272. *Power to collect* is an essential element of tax jurisdiction.

²⁵⁰ See Traynor, *supra* note 247, at 268.

²⁵¹ See Traynor, *supra* note 247.

²⁵² See Traynor, *supra* note 247. Former § 17980, renumbered today as § 18005, provides a credit for trustee taxes paid to its state of domicile, e.g. Missouri, against California taxes paid by the beneficiary on the same income. See Cal. Rev. & Tax. Code § 18005 (West 2004). Although the *McCulloch* opinion doesn't mention availability of the credit, § 18005 provides a credit to the beneficiary for taxes the trustee paid on the same income to Missouri, thereby properly minimizing taxation of the same income by two different states.

²⁵³ See Traynor, *supra* note 247.

²⁵⁴ See Traynor, *supra* note 247.

²⁵⁵ See Traynor, *supra* note 247.

²⁵⁶ See *supra* note 73.

principal and income statutes applicable to trusts.²⁵⁷

Professor Traynor's article explored ways in which such states could tax such accumulated income within the bounds of due process. Clearly, Traynor felt and stated that a state's "jurisdiction to tax should be found wherever substantial claims to tax are reinforced by effective power to compel payment."²⁵⁸ "Substantial claims to tax" refers to the state's due process nexus to the income. "Effective power to compel payment" refers to *in personam* jurisdiction over the taxpayer in order to compel his tax payment. "The power to tax is empty without the power to collect the tax; hence the power to collect is a necessary predicate to a state's power to tax."²⁵⁹ Traynor then analyzed the due process quandary of how to tax a foreign trustee's accumulated income when a taxing state lacked due process nexus with the trustee and current income was not being distributed to its resident beneficiary.²⁶⁰

His solution was that such a state should tax its own residents, whether settlor or beneficiary, *not* the foreign trustee who lacked nexus with the taxing state.²⁶¹

Traynor proposed taxing a resident settlor pursuant to Grantor trust principles, although grantor trust laws²⁶² were not enacted for another twenty years. Regarding beneficiary taxation, Traynor stated, "if a tax is effectively to reach such income, and thereby the recipient's ability to pay, *it must be imposed at the domicile of the beneficiary when the income is currently distributable.*"²⁶³ Traynor was proposing a system of

²⁵⁷ See Traynor, *supra* note 247.

²⁵⁸ See *supra* note 252.

²⁵⁹ See Lowy, *supra* note 54. "Power" to collect tax, i.e. personal jurisdiction over the taxpayer, pairs the jurisdiction to tax with *in rem* or *in personam* jurisdiction.

²⁶⁰ See Traynor, *supra* note 247.

²⁶¹ Clearly, Traynor was analyzing circumstances in which the trustee, settlor, and beneficiary lived in different states, and trust income was accumulated, not distributed. Traynor, *supra* note 247, at 274 ("[V]arious parties are domiciled in different states.").

²⁶² *Id.* at 284-91.

²⁶³ See Traynor, *supra* note 247, at 275 (emphasis added). Traynor opposed taxing the entire accumulated income to the California beneficiary solely in the year of distribution since progressive rate schedules would impose an inappropri-

beneficiary taxation upon distribution of accumulated income twenty years prior to enactment of the federal throwback rules which were later enacted based on his tax principles.²⁶⁴

Traynor recognized that the state of a beneficiary's residence was limited first, to taxation of its resident beneficiary, and second, to imposing such a tax only when the income was distributable.²⁶⁵ Traynor's position is absolutely clear and diametrically opposed to Connecticut taxing the New York trustee based solely on the domicile of the beneficiary. Traynor explained, "The state of either the Settlor's or the beneficiary's domicile would likewise have difficulty in collecting a tax from a non-resident trustee on income which was neither produced, received nor enjoyed within its borders during the period of accumulation"²⁶⁶

Traynor was also concerned about the validity under state law to tax the resident beneficiary on accumulated trust income even when distributed.²⁶⁷ The issue, as in *Simmon*, was whether state principal and income laws permitted taxation of the distribution as "income."²⁶⁸ His concern was based on the trust's accumulated income having blended with principal and having become a nontaxable gift, bequest, or return of capital when later distributed.²⁶⁹

ate tax burden on such beneficiary. See Traynor, *supra* note 247. Thus, following Traynor's philosophy, the beneficiary's tax was calculated separately for each year of accumulation under section 17745(d). See *infra* note 300.

²⁶⁴ The throwback rules of sections 665 through 668 taxed distributions of accumulated income by a complex trust roughly "as though they had been distributed to the beneficiaries in earlier years," with "credit for taxes paid by the trust when it reported the income. The throwback rules [were] intended to curb use of accumulation trusts as tax-avoidance devices." BORIS I. BITTKER & LAWRENCE LOKKEN, *FEDERAL TAXATION OF INCOME, ESTATES AND GIFTS*, ¶ 81.5.1 (2d. ed. 1990)

²⁶⁵ *Infra* note 268.

²⁶⁶ See Traynor, *supra* note 247, at 315. Again "power" to collect tax is an essential element of jurisdiction to tax. Traynor recognized there was no personal jurisdiction, i.e. no power, to collect the tax from a nonresident trustee.

²⁶⁷ There are constitutional "realization" issues in taxing a beneficiary prior to actual distribution, assuming the beneficiary cannot control distribution. If the beneficiary controls distribution, constructive receipt satisfies realization. See *Comm'r v. Glenshaw Glass*, 348 U.S. 426 (1935).

²⁶⁸ See Traynor, *supra* note 247.

²⁶⁹ "The validity of such a tax would depend largely upon the theories of the

The foregoing concerns, along with others,²⁷⁰ and the analysis of Professor Traynor were addressed and reflected in the unique California Statute²⁷¹ analyzed in *McCulloch*.²⁷² As Advisor to the California State Board of Equalization,²⁷³ it is likely Professor Traynor participated in drafting the unique California statutes modeled after his own proposals. First, Professor Traynor's analysis of the constitutional power "to compel payments"²⁷⁴ by the state of the settlor's and beneficiaries' residence resulted in a tax actually imposed on a California resident beneficiary.²⁷⁵ Likewise, the constitutional issue of realization²⁷⁶ resulted in California delaying taxation until income was actually distributed to the resident beneficiary. The state statutory issue of capital versus income²⁷⁷ was addressed by revising California's principal and income statutes.²⁷⁸ Double taxation criticism under the Commerce Clause was satisfied by allowing the beneficiary a tax credit for taxes paid by the trustee to its state of residence.²⁷⁹

The discussion below reflects that Traynor's theories were reflected and implemented in the California statutes and in the *McCulloch* opinion. The opinion noted that the tax was actually

states with regard to the nature of capital and income. Thus, in *Commissioner v. Simmon*, the Massachusetts Supreme Court held invalid an income tax regulation making accumulated income taxable to the resident beneficiary when received from a nonresident trustee on the ground that such accumulated income constituted 'capital' and not 'income' in the hands of the beneficiary." Traynor, *supra* note 247, at 281 (footnote omitted); see *Simmon*, discussed *supra* note 73.

²⁷⁰ Professor Traynor also raised the concern that taxation of a lump sum on distribution would throw income into higher brackets than imposition of an annual tax on smaller, annual income. See Traynor, *supra* note 247, at 280-81. Taxing income annually, as was done in *McCulloch*, is required by the California statute. See *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412, 417 (Cal. 1964).

²⁷¹ See *infra* note 296.

²⁷² See *supra* note 23.

²⁷³ See *supra* note 248 and *infra* note 310.

²⁷⁴ Absence of "power" to compel payment from the nonresident trustee equates to lack of in personam jurisdiction over such nonresident trustee.

²⁷⁵ See *infra* note 273; CAL. REV. & TAX. CODE § 17745(a) (West 2004).

²⁷⁶ See *Comm'r v. Glenshaw Glass*, 348 U.S. 426 (1935).

²⁷⁷ For discussion of *Simmon*, see *supra* note 73 and accompanying text.

²⁷⁸ See CAL. REV. & TAX. CODE § 17745(c) (West 2004).

²⁷⁹ See *id.* § 18005.

imposed on the California resident beneficiary, not the foreign trustee.²⁸⁰ The tax was calculated on each year's annual income, not on the accumulated lump sum actually distributed.²⁸¹ The beneficiary's obligation to pay was postponed until actual distribution.²⁸² The beneficiary absolutely did not pay the tax previously levied against the trustee.²⁸³

As stated in §17745(d), "The tax attributable to the *inclusion of that distributed income in the gross income of that beneficiary* for the year...had it been includable in the gross income of the beneficiary ratably ...for the period that the trust accumulated or acquired income for that contingent beneficiary . . .

"284

The *McCulloch* opinion and Professor Traynor's article stand for the clear proposition that a state of the settlor's or beneficiary's residence cannot constitutionally tax a foreign trustee. *Gavin's* citation of *McCulloch* as authority was wholly misplaced.

IX. *MCCULLOCH* IS NOT PRECEDENT FOR *GAVIN*.

*McCulloch*²⁸⁵ is the only case the *Gavin* opinion even attempted to claim as precedent for its holding as to Connecticut's taxation of the New York inter vivos trustee. *Gavin* mischaracterized *McCulloch* as "a closely analogous"²⁸⁶ case. *McCulloch* is not closely analogous to *Gavin* on its facts, on applicable law, or on its holding. As analyzed below, *McCulloch* in no way established even the slenderest precedential thread for *Gavin's* wholly unprecedented holding.

On the facts, *McCulloch* was an action to recover income taxes assessed against and previously paid by plaintiff, a California resident who was both a trustee²⁸⁷ and beneficiary of a

²⁸⁰ *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412, 419 (Cal. 1964).

²⁸¹ *Id.*

²⁸² *Id.*

²⁸³ *Id.*

²⁸⁴ Clearly, this statute was the predecessor and model for the federal throw-back rules. See *supra* note 265.

²⁸⁵ See *supra* note 23.

²⁸⁶ *Chase Manhattan Bank v. Gavin*, 733 A.2d 782, 802 (Conn. 1999).

²⁸⁷ See CAL. REV. & TAX. CODE § 17743 (West 1994) ("Where the taxability of

foreign trust.²⁸⁸ The California resident was one of three trustees.²⁸⁹ Pursuant to a very unique California statute,²⁹⁰ the beneficiary had paid the California tax levied on such resident beneficiary.²⁹¹ The California beneficiary's California tax was calculated by inclusion of income distributed from the nonresident trustee on the transferee beneficiary's personal tax return.²⁹²

McCulloch was more closely analogous to *Simmon*²⁹³ than *Gavin*, except that California had a statute²⁹⁴ which had no

income under this chapter depends on the residence of the fiduciary and there are two or more fiduciaries for the trust, the income taxable under Section 17742 shall be apportioned according to the number of fiduciaries resident in this state pursuant to rules and regulations prescribed by the Franchise Tax Board"). One-third was an allocable portion in *McCulloch*. See *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412, 415 (Cal. 1964).

²⁸⁸ *McCulloch*, 390 P.2d at 414-15. The testamentary trust was formed and administered under the laws of the state of Missouri under the will of a Missouri decedent. See *id.* at 415-16.

²⁸⁹ *Id.* at 414. In *Gavin*, no trustee was a resident of Connecticut. *Gavin*, 733 A.2d at 787.

²⁹⁰ CAL. REV. & TAX. CODE § 17745 (West 1994). See *infra* note 295. Connecticut had no similar statute and the California statute was essential to taxation of the California beneficiary in *McCulloch*. See *McCulloch*, 390 P.2d at 416.

²⁹¹ *McCulloch*, 390 P.2d at 415.

²⁹² *Id.* at 415. Section 17745(d) taxes the beneficiary anew on the distributed "income" and never makes the beneficiary liable as a transferee for the unpaid taxes of the foreign trustee. § 17745(d). See *infra* note 310.

²⁹³ *Comm'r of Corps. & Taxation v. Simmon*, 198 N.E. 741 (Mass. 1935).

²⁹⁴ Section 18102 (now § 17742) provided as follows:

Except as otherwise provided in Articles 2 and 4 of this chapter, the income of an estate or trust is taxable to the estate or trust [nevertheless, See Section 18106 below and *infra*, note ?]. The tax applies to the entire net income of an estate, if the decedent was a resident, regardless of the residence of the fiduciary or beneficiary, and to the entire net income of a trust, if the fiduciary or beneficiary is a resident, regardless of the residence of the settlor. § 18102 (repealed 1983).

Section 18103 read as follows:

Where the taxability of income of a trust depends on the residence of the fiduciary and there are two or more fiduciaries for the trust, the income taxable under Section 18102 shall be apportioned according to the number of fiduciaries resident in this State pursuant to rules and regulations prescribed by the commissioner. § 18103 (repealed 1983).
Section 18105 provided in part: "[t]axes on income of an estate or trust which is taxable to the estate or trust . . . shall be paid by the fiducia-

equivalent in Massachusetts, Connecticut, or anywhere else. The provisions currently in effect are found in sections 17742 through 17745 of the California Revenue and Taxation Code.²⁹⁵

Section 17745²⁹⁶ is entitled "Income [T]axable to [B]eneficiaries." The title "Income [T]axable to [B]eneficiaries" is clearly an irrefutable declaration that "[i]ncome" is taxable to the "beneficiar[y]." Section 17745 specifically provides as follows:

(a) If, for any reason, the taxes imposed on income of a trust which is taxable to the trust because the fiduciary or beneficiary is a resident of this state are not paid when due and remain unpaid when that income is distributable to the beneficiary, . . . if the taxes are not paid when due, *such INCOME shall be taxable to the beneficiary when distributable to him. . . .*

. . .

(d) The tax attributable to the *INCLUSION OF THAT INCOME IN THE GROSS INCOME OF THAT BENEFICIARY FOR THE YEAR . . .* shall be the aggregate of the taxes which would have been attributable to that income *had it been included in the gross income OF THAT BENEFICIARY* ratably for the year of distribution and the five preceding taxable years, or *for the period that the trust accumulated or acquired income for that contingent beneficiary, whichever period is the shorter.*²⁹⁷ (Emphasis added)

There are two reasons why the beneficiary is taxed on the distributed income but spread over the years the trust had realized such income. First and foremost, it is a mechanism preferred by Professor Traynor for reducing the tax penalty on

ry." § 18105 (repealed 1955).

Yet, when such tax was not paid by the trustee, section 18106 [renumbered § 17745] imposed personal liability for taxes on trust *income* on the resident beneficiary to include such income in the beneficiary's tax return when the income was distributed. See *McCulloch* 390 P.2d at 416-17; see *infra* note 297.

²⁹⁵ CAL. REV. & TAX. CODE §§ 17742-17745 (West 1994).

²⁹⁶ CAL. REV. & TAX. CODE § 17745 (West 1994).

²⁹⁷ *Id.* (emphasis added).

the beneficiary.²⁹⁸ The penalty arose from taxing five or more years of income accumulation to the beneficiary in a single year.²⁹⁹

Secondly, when a trust accumulates income, generally the trust is taxed and not the beneficiary. When the accumulated funds are later distributed to the beneficiary, federal and state tax law treat the income as having been inextricably commingled with and become a part of nontaxable trust principal.³⁰⁰ This tax principle "would prevent taxation of accumulated income when received by the beneficiary. This would in effect obstruct taxation of such income altogether whenever it were derived from intangibles, in view of the difficulties in reaching either of the other parties [settlor or trustee]."³⁰¹ Having a statute that levies, but does not collect, a tax on the beneficiary during each year the trust has income avoids a later argument by the beneficiary that he or she can't be taxed in the year of distribution since the previously accumulated income might now constitute principal under state law.³⁰²

This second purpose is also clearly reflected in section 17745, subsection (c), which appears to be a direct reaction to the holding in *Simmon*,³⁰³ as follows:

(c) The tax on that *income which is taxable to the beneficiary*

²⁹⁸ Traynor, *supra* note 247, at 279-85. Substantially the same approach was utilized in the federal throwback rules where a complex trust was used to shelter income. See *infra* note 265.

²⁹⁹ This statute deliberately reflects the sentiment expressed by Professor Roger Traynor in his classic tax article, *supra* note 249, that the trust income distributed should not be included in the aggregate on the beneficiary's tax return for the year of distribution. Due to progressive rate schedules, taxing the income from years of trust accumulation to the beneficiary entirely in the year of distribution would constitute a tax penalty on the beneficiary who exercised no control over the trust's decision to accumulate.

³⁰⁰ See discussion of *Simmon*, *supra* note 73 and accompanying text. See *infra* note 310.

³⁰¹ See Traynor, *supra* note 73.

³⁰² See *infra* note 283. This is a state principal and income statutory issue, not a constitutional question. The plaintiff/beneficiary in *McCulloch* unsuccessfully made this argument. *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412, 417 (Cal. 1964).

³⁰³ *Comm'r of Corps. & Taxation v. Simmon*, 198 N.E. 741 (Mass. 1935).

under subdivisions (a) or (b) *is a tax on the receipt of that income distributed or on the constructive receipt of that distributable income. For purposes of this section income accumulated by a trust continues to be income even though the trust provides that the income (ordinary or capital) shall become a part of the corpus.*³⁰⁴

The *McCulloch*³⁰⁵ opinion did only what the statute mandated, no more and no less. In *McCulloch*, no tax levied on the nonresident trustee was ever paid by anyone.³⁰⁶ No such tax was paid by the trustee. Nor was any such tax on the nonresident trustee ever paid by the California resident beneficiary. Rather, the California resident beneficiary paid only a separate and distinct tax calculated on the pass-through of the trust's income to and inclusion with the beneficiary's other taxable income and deductions.³⁰⁷ Eminent academic commentators clearly disagree.³⁰⁸

In *McCulloch*, a Missouri trustee earned, but accumulated, income each year from 1946 through 1950. "The trust paid state income taxes to Missouri upon its income for the years 1946 to 1950, inclusive; it paid no income tax to California during that period."³⁰⁹ California "imposed upon plaintiff

³⁰⁴ CAL. REV. & TAX. CODE § 17745(d).

³⁰⁵ *McCulloch v. Franchise Tax Bd.*, 390 P.2d 412 (Cal. 1964).

³⁰⁶ *Id.* at 416. See *infra* note 311 and accompanying text.

³⁰⁷ See *McCulloch*, 390 P.2d at 415-16.

³⁰⁸ 2 JEROME R. HELLERSTEIN & WALTER HELLERSTEIN, *STATE TAXATION* ¶ 20.09 (3d ed. 2000), states as follows:

Finally, the California Supreme Court relied on the presence of an in-state beneficiary to sustain the state's power to tax a trust, created by a nonresident decedent and administered outside the state, on income it was accumulating for the in-state beneficiary. The trust, which was administered in Missouri, had failed to pay taxes to California on the accumulated income, and, upon distribution of the accumulated income to the beneficiary, the beneficiary became liable for the *trust's* tax.

Again, the beneficiary did not "bec[ome] liable for the *trust's* tax" as it would have if the statute had actually provided for a "transferee tax." Instead, the beneficiary became liable for the beneficiary's own tax on income passed through from the trust.

³⁰⁹ *McCulloch*, 390 P.2d at 416. "Since the trust failed to pay this state's tax upon its annual income, *California can constitutionally tax the beneficiary* at the

[beneficiary] liability for income taxes upon the accumulated income [subsequently] distributed by the trust ³¹⁰

In 1951, when the terminal distribution was made to the California beneficiary, California, then and only then, levied a separate income tax on the beneficiary.³¹¹ The tax calculation and its collection clearly and deliberately fell solely on the California beneficiary, *not* the Missouri trustee.³¹²

Although the statute ostensibly undertook initially to tax the foreign trustee on its undistributed annual income in the years 1946 through 1950, both the statute and the Supreme Court of California ultimately placed the real tax burden on the California beneficiary.³¹³

The California Supreme Court in its *McCulloch* opinion clearly and correctly held that the residence of a beneficiary of a foreign trust was a sufficient basis for California to tax such California resident beneficiary.³¹⁴ As noted by the Connecticut Supreme Court in *Gavin*, “[w]e agree with the reasoning of the California Supreme Court that ‘[t]he laws of the state of resi-

time he receives the accumulated income” *Id.* at 414-15 (emphasis added).

³¹⁰ *Id.* at 415. The statute was the predecessor of the federal “throwback rules” imposing a tax directly on the beneficiary at the time of distribution. *See, e.g.*, I.R.C. § 665 (1954). California’s very unique tax on the beneficiary of a trust is not a typical transferee tax provision. Indeed, it is a misnomer to describe it as a “transferee tax” at all. In actuality, it is merely a new direct tax on the trust beneficiary as the result of trust income being distributed, i.e., “transferred to,” and taxed directly to such beneficiary. For typical transferee tax liability, *see*, for example, I.R.C. § 6901 (2000) (Transferee of property is liable for Transferor’s unpaid income and/or transfer taxes plus a pro rata portion of any interest owed by Transferor thereon to the extent of fair market value of property received from the Transferor.)

³¹¹ *McCulloch*, 390 P.2d at 415. *See* CAL. REV. & TAX. CODE, § 17745(a) (West 1994) (“such income shall be taxable to the beneficiary when distributable to him. . . .”).

³¹² *See* CAL. REV. & TAX. CODE § 17745 (West 1999). *See supra* note 313 and accompanying text.

³¹³ *McCulloch*, 390 P.2d at 419-21. The opinion did not require the beneficiary to pay the trust’s tax or associated interest for late payments back to 1946. *See id.* at 421.

³¹⁴ *Cf. Maguire v. Trefry*, 253 U.S. 12 (1920) (holding that Massachusetts can tax its resident beneficiary on distributed current income from a Pennsylvania trust). *See* Walter L. Nossaman, *State Taxation of Income*, 24 CAL. L. REV. 524 (1936).

dence afford benefit and protection to the *resident beneficiary* [no reference to *the foreign fiduciary*]"³¹⁵ Further, the *Gavin* opinion also correctly agreed with the California Supreme Court's statement in *McCulloch* that:

The tax imposed by California upon the *beneficiary* [no reference to the foreign trustee] is constitutionally supported by a sufficient connection with, and protection afforded to, [the plaintiff as such *beneficiary* California grants the *beneficiary* [no reference to the foreign trustee] the interim protection of its laws so that he [the beneficiary] may ultimately obtain the benefit of the accumulated income [The] [p]laintiff's [beneficiary's] residence here confers the essential minimum connection . . . necessary for due process of law [i.e., unquestionably satisfies due process for taxation of the resident California or Connecticut beneficiary, but does *not* support taxation of the foreign trustee in either case].

The *Gavin* opinion then jumps from Connecticut's right to tax its own resident beneficiary as in *McCulloch*, to claim a state's right to tax a foreign trustee who has no connection, minimum or otherwise, with the State of Connecticut.³¹⁶ This is an unsustainable and utterly indefensible leap in logic.

This flawed logic then led the Supreme Court of Connecticut to make its decision to disregard a contradictory, but well-reasoned, body of precedent. As the *Gavin* opinion itself stated in footnote 25, "[w]e disagree, therefore, with the reasoning of those cases relied on by the plaintiff in which courts have found the domicile of a beneficiary of an inter vivos trust insufficient for due process purposes."³¹⁷ Courts have long held

³¹⁵ Chase Manhattan Bank v. Gavin, 733 A.2d 782, 802 (Conn. 1999) (quoting *McCulloch*, 390 P.2d at 418-19) (emphasis added).

³¹⁶ *Gavin*, 733 A.2d at 803, 806. The mere attributed benefit of Connecticut law is insufficient for either due process or commerce clause purposes. See *supra* note 310. Of course, a beneficiary's residence in a state supports taxation of such resident beneficiary. This is all the California Supreme Court did in *McCulloch*.

³¹⁷ *Gavin*, 733 A.2d at 802 n.25. The court refers to *Blue v. Dep't of Treasury*, 462 N.W.2d 762 (Mich. Ct. App. 1990); *Potter v. Taxation Div. Dir.*, 5 N.J. Tax 399 (N.J. Tax Ct. 1983); *Mercantile-Safe Deposit & Trust Co. v. Murphy*, 203 N.E.2d 490 (N.Y. App. Div. 1964). *Id.*

that the domicile of a beneficiary was an insufficient contact to satisfy due process requirements for taxation of an otherwise foreign trustee.³¹⁸

The *McCulloch* opinion clearly addressed "due process" from the standpoint of California imposing a tax on its own resident beneficiary just as Missouri had properly taxed the trustee resident in Missouri:

We find no reason in constitutional principles or in practical application . . . why the tax [on the beneficiary] founded upon the residence of the beneficiary should not be sustained. We shall point out that just as the protection and benefits afforded by the state of the residence of the trustee [e.g., Missouri]³¹⁹ serve as the basis for the constitutionality of the [Missouri] tax as to him, so do those factors serve as the basis for the constitutionality of the [California] tax as to the beneficiary.

In *McCulloch*, Missouri taxed its trustee.³²⁰ In like manner, California imposed a tax on its resident California beneficiary.³²¹ The opinion, like § 18106,³²² repeatedly made it clear that California's tax was ultimately on the beneficiary:

We conclude that California could constitutionally tax plaintiff as the resident beneficiary upon the accumulated income when it was distributed to him.

Connecticut in *Gavin* and Massachusetts in *Simmon*, unlike California in *McCulloch*, lacked state statutes to impose a tax on the Connecticut or Massachusetts beneficiaries, respectively, upon receipt of the trust's accumulated income.

The *Gavin* opinion engaged in a logical disconnect by hold-

³¹⁸ See *supra* Part III. B.; *Simmon*, 198 N.E. 741.

³¹⁹ With one of three trustees resident in California, California could have imposed a pro rata one-third tax based on trustee residence in California. See *supra* note 295 regarding former § 18103.

³²⁰ *Id.* at 415.

³²¹ *Id.* at 416.

³²² Section 18106 has been renumbered as § 17745. See *supra* note 297 and accompanying text of subsections (a) and (d).

ing that since Connecticut could,³²³ but statutorily did not, tax the Connecticut beneficiary, it had nexus to tax the foreign trustee.³²⁴ Therein lies the fundamental and very dangerous flaw in the opinion's unprecedented holding.³²⁵

X. PRE- AND POST-GAVIN STATE RESIDENCY RULES FOR TRUSTEES

States have historically applied myriad nexus rules to trustee taxation.³²⁶ California's approach, established statutorily and judicially upheld by Professor Roger John Traynor, was discussed above.³²⁷ Idaho, similar to Missouri in *Swift*,³²⁸ taxes "trusts" as residents if they satisfy at least three out of five (six in *Swift*) characteristics.³²⁹

A New York "resident" trust, i.e., created by a New York resident settlor, is a nonresident trust for income tax purposes if:

- (1) all the trustees are domiciled in a state other than New York State;³³⁰ (2) the entire corpus of the trust, including real and tangible property is located outside of New York State [non-business intangibles deemed located outside the state]; and (3) all income and gains of the trust are derived or

³²³ Under Due Process, Connecticut *could* have taxed its resident beneficiary upon distribution. Under Connecticut state statutes, however, Connecticut, like Massachusetts in *Simmon*, could *not* tax the resident on the distribution of accumulated income.

³²⁴ *Gavin*, 733 A.2d at 793.

³²⁵ Attribution to the nonresident of benefits of Connecticut law to the Connecticut beneficiary was rejected by the Supreme Court in *Shaffer*, since the necessary relationship to the fiduciary's income earned in New York was insufficient for both due process (*Shaffer*) and commerce clause (*Complete Auto*, Step Four) purposes. See *supra* note 310.

³²⁶ Nenno and Sparks, *Delaware Dynasty Trusts, Total Return Trusts, and Asset Protection Trusts*, printed by Wilmington Trust Bank.

³²⁷ See discussion *supra* Part V.

³²⁸ *Swift v. Dir. of Revenue*, 727 S.W.2d 880 (Mo. 1987) (en banc).

³²⁹ See *supra* note 145.

³³⁰ See, e.g., *In re John Frankel Trust*, 1980 N.Y. Tax LEXIS 39 (New York State Tax Commission) (holding that a trust created by New York resident and administered under Connecticut law could be taxed as "resident" New York trust where majority of trustees were New York residents).

connected from sources outside of New York State, determined as if the trust were a nonresident.”³³¹

New Jersey has a similar standard: “If a resident trust or estate does not have any assets in New Jersey or income from New Jersey sources, and does not have any trustees or executors in New Jersey, it is not subject to New Jersey tax.”³³²

At least one nationally respected Connecticut commentator, Mr. Frank Berall,³³³ at the time of the adoption of Connecticut’s current income tax on trusts, analyzed Connecticut’s tax statute as being in the mainstream.³³⁴ In many respects, Mr. Berall stated parallels between New York and Connecticut law as follows:

A resident trust will *not be taxed* if all its trustees are nonresidents, its entire corpus is located outside Connecticut and all its income and gains are derived from or connected with sources outside Connecticut.³³⁵

Anticipating the *Gavin* issue, though not its holding, Mr. Berall further stated:

The fact that such a trust still has resident beneficiaries should *not* make it a resident trust. *A tax cannot be imposed by Connecticut merely because trust beneficiaries reside here, if the trust’s assets are possessed by a trustee in another state; otherwise it would violate the due process provision of the Fourteenth Amendment to the federal constitution.*³³⁶

³³¹ N.Y. COMP. CODES R. & REGS. tit. 20, § 105.23 (2003).

³³² Form NJ-1041, Instructions, 2005 New Jersey Gross Income Tax Fiduciary Return, http://www.state.nj.us/treasury/taxation/pdf/other_forms/tgiee/2005/051041i.pdf.

³³³ Frank S. Berall, formerly of the Hartford Bar and former frequent national lecturer on income taxation of trusts and estate planning.

³³⁴ Frank S. Berall & Suzanne Brown Walsh, *Income Taxation of Estates and Trusts*, 65 CONN. B.J. 377 (1991).

³³⁵ *Id.* at 378 (emphasis added).

³³⁶ *Id.* (citing *Safe Deposit & Trust Co. v. Virginia*, 280 U.S. 83 (1929) (emphasis added); *Guaranty Trust Co. v. Virginia*, 305 U.S. 19, 23 (1938); *Mercantile Safe-Deposit and Trust Co. v. State Tax Commission*, 19 A.D.2d 765 (N.Y. App. Div. 1963). These cases were rejected by the *Gavin* opinion’s wholly unprecedented holding. See *supra* note 286 and accompanying text.

It seems that local politics and desperation for more state revenue have contributed to the dangerous, unprecedented holding of *Gavin*.³³⁷

CONCLUSION

The result in *Gavin* is not consistent with the relations of states within the American Union or with the expectations of U.S. citizens.³³⁸ Instead of butchering constitutional limitations on state taxation of foreign trustees, whether testamentary or *inter vivos*, states should merely review and revise their own rules on taxation of grantors, trustees, and beneficiaries.

Unconstitutionally expansive Founder State Trust rules are not the answer. Like Justice Traynor, states should look to adoption and/or modification of traditional federal grantor trust

³³⁷ The same pressures can be seen in Alabama's new resident trust tax statute. HB19 was part of a comprehensive trust tax reform package enacted by the Alabama legislature. HB19 defined a resident trust as a trust that is described in either (a) or (b) below:

(a) A trust: (1) Created by the will of a decedent who was an Alabama resident at death or . . . who was an Alabama resident [including seven-month nondomiciliary resident] at the time the trust became irrevocable; and (2) For more than seven months during the taxable year, an individual who either resides in [nondomiciliary resident] or is domiciled in Alabama is either a *fiduciary* of the trust or a *beneficiary* to whom distributions may currently be made. [OR] (b) a trust that has Alabama as its principal place of administration during the taxable year.

Clearly, a nondomiciliary resident could meet the requirements of both (a)(1) and (a)(2). A Florida domiciliary could, for example, have established a Florida or a Delaware trust and be a permissive beneficiary of such a trust. A clear example could be a Delaware self-funded spendthrift trust. If, for unrelated business reasons, the Florida resident spent seven months in Alabama and either died in an auto crash or for some other reason the trust became irrevocable; the Delaware trust of a Florida resident with a Delaware trustee is forever an Alabama resident trust.

Tying taxation solely to trust administration is similarly flawed. A Texas resident could establish a trust with the Texas affiliate of an Alabama bank holding company. If, wholly unbeknownst to the Texas settlor, all the holding company's trust administration is conducted at its Alabama headquarters, the trust will be taxed forever as a resident Alabama trust.

³³⁸ See Jacob, *supra* note 5.

and throwback rules. Too many states merely conform their grantor trust rules to federal grantor trust rules. Under I.R.C. § 679, the federal government taxes a U.S. citizen on all income of a foreign, i.e., outside the U.S., trust that has a U.S. beneficiary and was established by the U.S. grantor. A state which conformed to federal grantor trust rules, could, under its counterpart of § 679, tax its citizen grantors on trusts they might establish in Bermuda as a tax haven. However, if the same state resident grantor created a similar trust in Delaware as a state tax haven, it is not a "grantor" trust for federal purposes or for conforming state law purposes.

States like Connecticut that are interested in taxing income placed in out-of-state trusts by Connecticut residents, should establish grantor trust rules which parallel § 679, but which define a "foreign" trust as being established and domiciled outside the state of Connecticut, not merely outside the U.S. This approach does, of course, subject the trust's income to the risk of double taxation, which should be mitigated by income allocation or tax credits as required by *Complete Auto's* commerce clause analysis.

Likewise, states like Connecticut, that want to tax a foreign trustee's accumulated income based on the state's benefits provided to the trust's beneficiary, should adopt laws similar to California's throwback rules. Such laws could parallel federal throwback rules which were in turn based on California's Traynor tax model. Throwback rules would impose the state income tax on its resident beneficiary at the time of a subsequent distribution. Today, under the U.S. Internal Revenue Code, federal "throwback rules" apply primarily to foreign, i.e. non-U.S., trusts.³³⁹ The federal rationale and need of the throwback rules for foreign trust taxation serves essentially the same remedial purpose states need to apply to out-of-state trusts with resident beneficiaries.³⁴⁰

Gavin is a badly flawed ruling which, in most respects, has no precedent whatsoever. It was founded on state desperation

³³⁹ I.R.C. § 679.

³⁴⁰ *Id.*

for revenues and local politics, reflecting the tax adage "Don't tax you, don't tax me, tax the fella behind the tree." In *Gavin*, the "you" and the "me" are Connecticut resident settlors and beneficiaries, and the "fella behind the tree" is a nonresident trustee.

A PROLEGOMENA TO FEDERAL STATUTORY INTERPRETATION: IDENTIFYING THE SOURCES OF INTERPRETIVE PROBLEMS

*J. Gordon Christy**

INTRODUCTION

The Constitution allocates sole legislative authority to Congress and ultimate interpretive powers to the federal courts.¹ How to harmonize these powers has remained unclear for almost 220 years.² If Congress, the legislature, has the sole authority to legislate, and if the results of legislating are evidenced in statutes, written texts, approved by the legislature, may the courts, in the interpretive process, constitutionally add to or subtract from the substantive or remedial effect of the written provisions of those statutes, including

* J.D. University of Texas at Austin. Associate Professor of Law and Director of Legal Analysis and Reasoning, Mississippi College School of Law. Professor Christy was previously Special Counsel to the Deputy General Counsel of Ford Motor Company, a partner in the Dallas, Texas, law firms of Johnson & Gibbs and Strasburger & Price, and Associate Professor of Law in the University of Oklahoma College of Law.

¹ U.S. CONST. art. I, § 1; U.S. CONST. art. III, §§ 1, 2. Both the executive and the various administrative agencies must interpret laws in the process of administering them. They, therefore, of necessity have interpretive power. That power is simply not the ultimate power. The final say resides with the federal courts.

² If one traces the doctrine of legislative supremacy in the Anglo-American tradition back to circa 1215, which is not entirely unreasonable, then harmonizing these powers has eluded jurists in the Anglo-American tradition for nearly 800 years. To put matters in an even longer-term perspective, the first known code of laws, the Sumerian Code of Lipit-Ishtar, was promulgated circa 2500 B.C. Francis Rue Steele, *The Code of Lipit-Ishtar*, 52 AM. J. ARCHAEOLOGY 425 (1948). The need to distinguish the powers of the legislator from those of the interpreter of statutes apparently extends back over 4,500 years.

express statutory definitions? In the absence of express statutory definitions, may the courts depart from the common, ordinary, or usual, including in special fields, technical meanings (collectively, as appropriate, "common meanings") of the words and phrases the legislature uses in statutes?³ Would not such additions, subtractions, and departures constitute legislating, a function reserved exclusively to the legislature? If courts, in the interpretive process, may constitutionally depart from statutory text, including express definitions and common meanings, what limits their authority to depart? There must be limits or there is no legislative supremacy. If, in the absence of clarifying theory, courts must, purely as a practical matter, on occasion depart from text and the common meanings of words and phrases, what prevents courts from abusing the interpretive power and becoming the ultimate legislators as well?

Until jurisprudence answers the foregoing questions, we will continue to have incredibly embarrassing lacunae of the most fundamental types in both constitutional law and legal theory. First, and most obviously, after operating under the Constitution for almost two hundred and twenty years, we still do not know how to give effect to the constitutional provision for legislative supremacy. Second, since we have no way of saying when statutes are interpreted properly, the law is a profession that, given the predominate role of statutory law in contemporary society,⁴ in significant part cannot justify the

³ With respect to any statute, the legislature clearly had the power to add to or subtract from the substantive and remedial provisions of the legislation enacted but did not do so. If the legislature did not exercise that power, can the courts constitutionally do so? The legislature, likewise, clearly had the power to provide special definitions for words and phrases used in statutes. If it did so, may the courts constitutionally depart from the legislatively assigned meanings of those words and phrases? If, on the other hand, the legislature selects given words and phrases to express the law and does not give them special meanings, i.e., does not provide special definitions, it would appear that the legislature had no intention of conferring any meanings but the common meanings. In light of that presumption, may the courts, in the interpretive process, legitimately depart from the common, standard, or usual meanings of those words?

⁴ ANTONIN SCALIA, *A MATTER OF INTERPRETATION* 13-14 (1997).

results it produces.

There have been numerous attempts to answer the questions above. Although intense intellectual effort ordinarily results in substantial progress, that has not been the case with respect to statutory interpretation. Rather than improving, our understanding of statutory interpretation has deteriorated over time. Proposed answers have proliferated, especially in recent years,⁵ and consensus has proved ever more elusive. There are currently at least fourteen theories of statutory interpretation in the United States alone and no telling how many variants of each.⁶ In spite of the number theories available, or maybe because of it, Justice Scalia has candidly admitted: "... American judges have no intelligible theory of what we do most,"⁷ which is interpret statutes. One scholar, in a highly thoughtful article, has even been led to set forth what he believes to be grounds for "*reject[ing]* the notion that there is a 'right' way"⁸ to interpret statutes and has concluded that we simply "*cannot* resolve the debate."⁹ This scholar believes that the search for the correct interpretive regime must give way to an explanation of "the seemingly endless struggles to determine the appropriate judicial approach toward legislative language."¹⁰

In reflecting on the chaotic state of the jurisprudence of

⁵ As William N. Eskridge, Jr., notes, since the early 1980s "theories of statutory interpretation have blossomed like dandelions in spring." WILLIAM N. ESKRIDGE, JR., *DYNAMIC STATUTORY INTERPRETATION* 1 (1994).

⁶ William N. Eskridge, Jr., lists in the neighborhood of thirteen current theories or schools of statutory interpretation, *i.e.*, intentionalism, purposivism, textualism, a group under the heading of "jurisprudential theories" (including liberal, legal process, coherence, and normativist theories, feminist republicanism, deconstruction, social construction, and critical pragmatism), legal realism, and his own theory of dynamic statutory interpretation. See WILLIAM N. ESKRIDGE, JR., *DYNAMIC STATUTORY INTERPRETATION* (1994). If we add in the observations of the critical legal studies movement, the number is pushed up to fourteen.

⁷ SCALIA *supra* note 4, at 14.

⁸ Timothy P. Terrell, *Statutory Epistemology: Mapping the Interpretation Debate*, 53 EMORY L.J. 523, 558 (2004) (emphasis in original). We might even classify Professor Terrell's views as a fifteenth theory of statutory interpretation.

⁹ *Id.* at 525 (emphasis in original).

¹⁰ *Id.* at 524.

statutory interpretation, the author was reminded of two facts. First, it is well known in cognitive science that solving a problem is often a function of first framing it properly. As the noted epistemologist Alvin Goldman observes: "The importance of how a problem is 'represented', or 'framed', cannot be underestimated."¹¹ "One conceptualization of a problem can make it hard to solve, while another makes it [relatively] easy."¹² In addition, failure to frame a problem properly can be the result of an outright mis-identification of the problem. Mis-identification means that proposed solutions will not be solutions to the actual problem. Further, with mis-identification, a solution will not merely be difficult, it is likely to be impossible. Second, it is recognized in science that a plethora of competing but inadequate theories almost always signals an underlying failure to properly frame a problem or a mis-identification of the problem.¹³

¹¹ ALVIN I. GOLDMAN, *EPISTEMOLOGY AND COGNITION* 132 (1986).

¹² *Id.* Goldman illustrates this point with the following example from Michael Posner:

Two train stations are 50 miles apart. At 2:00 P.M. two trains start toward each other, one from each station. Just as the trains pull out of the stations, a bird springs into the air in front of the first train and flies ahead to the front of the second. When the bird reaches the second, it turns back and flies toward the first. It continues to do this until the trains meet. If both trains travel at 25 miles per hour and the bird flies at 100 miles per hour, how many miles will the bird have flown before the trains meet?

Id. (rephrasing MICHAEL POSNER, *COGNITION: AN INTRODUCTION* 150-51 (Scott, Foresman 1973)). Goldman points out:

If the problem is conceptualized in terms of the bird's *trajectory*, one must calculate how far the bird goes on each trip between the trains. The problem is then difficult. But if it is conceptualized in terms of flight *time*, it becomes trivial. The trains must travel one hour before meeting (50 miles at a closing speed of 50 miles per hour), and at 100 miles per hour the bird will fly 100 miles.

Id. (emphasis in original).

¹³ See THOMAS KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* (3d ed. 1996). As Steven Pinker observes: "[I]f some phenomenon seems to make no sense no matter how we look at it, we are probably in ignorance of deep and far-ranging principles." STEVEN PINKER, *LEARNABILITY AND COGNITION* 1 (1989). By "deep," Professor Pinker means "below the surface," not necessarily "abstrusely remote."

Those facts led the author to hypothesize that the current conundrum regarding statutory interpretation must be due to a failure to adequately conceptualize and, possibly, even a mis-identification of the problems encountered in the interpretation of statutes. Since jurists generally focus solely on the problems that statutory rule systems present *after* they are formulated, the author began to suspect that it might prove useful to identify the problems we face in *formulating* statutory rule systems. The hypothesis that emerged was that only by first identifying in detail the problems faced in formulating statutory rule systems can we truly understand the problems encountered in interpreting statutes and, thereby, discover, or at least have a decent chance of discovering, an effective solution.

This article explains in detail the problems we face in formulating statutory rule systems. By way of introduction, the problems are that it is impossible *ex ante* to fully accomplish three essential tasks. First, it is *ex ante* often impossible to fully formulate and virtually always impossible fully to refine the purposes of a statutory rule system that addresses a complex underlying subject matter. Second, if the underlying subject matter is at all complex, it is *ex ante* impossible to

As Mark C. Baker notes:

The key to resolving a paradox often lies in the imagination. In a paradox where some experience points to one conclusion and other experience seems to point to its opposite, what is needed is not simply more experience. . . . [W]hat is needed is some new idea that can widen the space of hypotheses. Then the conflicting evidence that seemed to lead to contradictory conclusions can be seen to converge on this new possibility.

MARK C. BAKER, *THE ATOMS OF LANGUAGE* 19 (2001). It is worth noting how simple many significant reconceptualizations have been. Nicolaus Copernicus' great insight was simply that things might be the other way around, i.e., that the planets revolve around the sun. Johannes Kepler realized that planets could travel in elliptical orbits. Einstein's theory of relativity in part grew out of the realization that since light travels at a finite speed, the way that different observers in the universe will order events will necessarily differ. And the geometers Lobatchewski, Helmholtz, and others realized that entire new geometries could be constructed by assuming that all lines that in Euclidean geometry are parallel meet at infinity or diverge at infinity. J. ALBERTO COFFA, *THE SEMANTIC TRADITION FROM KANT TO CARNAP: TO THE VIENNA STATION* 41-61 (1991).

formulate a statutory rule system that is substantively and remedially complete in light of its underlying purposes and other purposes simultaneously operative in the law. And third, it is *ex ante* impossible to ensure that the meanings of the words and phrases that comprise a statutory rule system are in all respects appropriate in light of those purposes. The three foregoing problems give rise to at least twenty-nine distinct sub-problems. The collective effect of these problems is that we regularly encounter situations that the parties who drafted a statute overlooked or failed to anticipate and, therefore, did not address. When we do encounter them, applying a statute in accordance with what may be called its "presumptive semantic content"¹⁴ will all too often produce results that are not only legislatively unintended but random, accidental, arbitrary, and also, in varying degrees, irrational, unfair, or unjust.

Identification of the problems makes it clear that the above hypothesis is, indeed, correct. Jurists, who generally attribute the law related problems faced in the interpretation of statutes¹⁵ to an ambiguous ambiguity,¹⁶ have, in large

¹⁴ The "presumptive semantic content" of a statute is the information encoded in the text on the basis of the words, phrases, and sentences employed, special definitions provided, and, in the absence of special definitions, the common meanings of the words and phrases employed, without regard to the larger context out of which the statute emerges. The presumptive semantic content of a statute may be divided into the "presumptive prescriptive content," the rules, viewed as commands to courts, evidenced in the sentences of the text, and the "presumptive lexical content," the common meanings of words and phrases employed and special statutory definitions.

¹⁵ There are also linguistic problems associated with the interpretation of statutes. These problems result from an additional, equally important, failing in conceptualization. Jurists, like almost everyone, with the exception of cognitive scientists and linguists (the term now used not only for individuals who have mastered many languages but generally for scholars and researchers in the field of linguistics), do not have adequate conceptualizations of language and linguistic meaning, communication, and comprehension. That problem is only briefly addressed in this article (see *infra* note 17), but is fully addressed in a companion article.

¹⁶ For example, the most exhaustive treatment of statutory interpretation on the American scene is Norman J. Singer's eight volume *STATUTES AND STATUTORY CONSTRUCTION*. NORMAN J. SINGER, *STATUTES AND STATUTORY CONSTRUCTION* (6th

part, both mis-identified and failed to properly frame the problems. The fundamental problems faced in connection with statutory interpretation are practical and are only most rarely caused by ambiguity. The solutions, accordingly, are not fundamentally linguistic. Since, as a society, we want our statutory rule systems to be purposively, substantively, and remedially complete (that is, effective) and to produce intended, consistent, and non-arbitrary (that is, rational and, on appropriate occasions, fair and just) results, interpreting a statute consists of determining when a statute should be given effect in accordance with its presumptive semantic content because the matters at issue were clearly to most likely within the contemplation of the parties constructing the rule system and when, in contrast, the matters at issue were overlooked or not anticipated and proper application of a statute requires the much more complex, extensive, indeed, unending tasks of (1)

ed. 2002). Based on a review of the literature, Professor Singer's treatise mentions ambiguity as the only source of the interpretive problems relating to statutes. As if that were not enough, juristic lore does not clearly distinguish between the types of ambiguity. There are, in fact, at least four primary phenomena that are denoted by "ambiguity." The most commonly denoted is conceptual, lexical ambiguity. Occasionally, vagueness is mentioned in connection with ambiguity, although seldom is there an attempt to understand vagueness in depth. (An exception is TIMOTHY A. O. ENDICOTT, *VAGUENESS IN LAW* (2000).) In any event, as Part IV of this article explains, there is a second form of non-conceptual, lexical ambiguity that flows from the distinction between prototypical and conceptually identical but non-typical uses of words. See *infra* Part IV. A third type of ambiguity has nothing to do with the meanings of words and noun and verb phrases per se but rather with the role that the denotatum or denotata of a word plays or play in a sentence or phrase. This is known as "syntactical ambiguity." Syntactical ambiguity occurs, for example in indirect discourse. In indirect discourse it is impossible to tell syntactically what is subject and what is direct object. Additional forms of syntactical ambiguity result from sentence constructions that do not make clear the scope of modifiers, i.e., adjectives and adverbs, or the particular referents of pronouns (anaphoric ambiguity). For a sophisticated discussion of syntactical ambiguity see LAWRENCE M. SOLAN, *THE LANGUAGE OF JUDGES* 28-92 (1993). And finally there is a type of ambiguity that results when one says or writes something that appears to conflict with something else one holds to be true or claims to be an objective. In this instance the ambiguity is not linguistic at all. Rather the ambiguity relates to the beliefs or intentions and purposes of the communicator. The question in this latter case is which belief or intentions and purposes actually hold, or can the apparent inconsistencies somehow be harmonized?

clarifying and refining its underlying purposes, (2) completing inevitably incomplete statutory rule systems in light of their underlying purposes and other purposes simultaneously operative in the law, and (3) modifying the meanings of words and phrases used in statutory rules to bring those meanings into line with the foregoing purposes.¹⁷

¹⁷ This may appear to be a quite odd definition of "interpretation," but it is, in fact, completely in line with the understanding of linguistic interpretation that contemporary linguistics has established. The process of linguistic interpretation is far more complicated and sophisticated than for millennia realized.

For millennia we have erroneously tended to view language as operating solely as a code. This view was first given philosophical respectability by St. Augustine in the 400s. ANTHONY KENNY, WITTGENSTEIN 153-54 (1973). For a discussion of the code theory of language see, for example, C. K. OGDEN & I. A. RICHARDS, *THE MEANING OF MEANING* 10-13 (1923); JOHN LYONS, 1 SEMANTICS 32-41 (1977). It is obvious that we do, in fact, use nouns and verbs to represent, e.g., things and actions; adjectives and adverbs to represent, e.g., qualities of things and actions; prepositions to represent, e.g., relations (e.g., "on" and "over") and direction (e.g., "to" and "away"); phrases to represent, e.g., the things that nouns and verbs represent as modified by the qualities that adjectives and adverbs represent; syntax (in English, position in a sentence) to specify, e.g., subject and direct object; and grammar to represent, e.g., time ("ed" indicates the past), gender, and number ("-s" represents the plural). Under the code view of language, that is fundamentally all there is to language. All linguistically communicated meaning (information) is encoded in meaningful correlates of words, phrases, sentences, syntax, and grammar. By definition, a communicator encodes in language all, and only, the information the communicator wants the recipients of his or her communication to comprehend. Accordingly, on this view, all that recipients of linguistic communications can properly comprehend is the information they obtain by decoding. This view of language, which is at the foundation of textualism, renders it impossible to understand how the legislature could possibly mean more than or something other than the information derivable by decoding statutory text.

The code view of linguistic communication is now philosophically and scientifically known to be false. That is because it is seriously incomplete. In using language to communicate, we do much more than simply encode and decode. The linguistic discipline of pragmatics has revealed that, in accordance with the demands of epistemology, linguistic communicators intend and expect the recipients of their linguistic communications to comprehend, and the recipients may and should comprehend, all the information that is encoded in meaningful correlates of words, phrases, sentences, syntax, and grammar *plus all the additional information that the recipients can properly infer from the combination of the linguistically encoded information and all other accessible, relevant information, including in particular, all accessible, relevant information regarding the communicator's intentions and purposes*. See, for example, the works cited *infra* notes 29-30. That is, linguistic interpretation, or comprehension, includes *both* the obvious encoding

Given the constitutional provision for legislative supremacy, the process for accomplishing those tasks would ideally always produce objective, politically and ideologically neutral, determinate, and undoubtedly legislatively intended results. The conclusion to this article briefly introduces the interpretive process that the insights above entail.¹⁸ Although that process cannot achieve the ideal, to borrow conceptually from the differential calculus, it can approach the ideal *qua* limit. If we recognize that results we must accept because the probabilities favor them, even if not certain, are determinate, we can in the vast majority of real world situations achieve highly determinate, objective, politically and ideologically neutral results that can be viewed as certainly to most likely intended by the legislature. And when we encounter situations in which the results are unquestionably not determinate, we can, nonetheless, provide justified outcomes, even if such outcomes may on occasion have some unavoidable political or ideological coloring. Since these situations tend to be at the fringes rather than the core of statutory rule systems, that is a quite acceptable, if not perfect, state of affairs.

and decoding process as well as a sophisticated process of inference drawing that uses as premises all relevant information, including information that may first become available to or be recognized by either the communicator or recipient well after a linguistic communication is uttered or written. The result is that linguistic communicators always mean more than they say or write, meaning, or information, that is derivable by inference. In that regard, statutes are like all linguistic communications. There is absolutely nothing odd or unusual about drawing inferences based all relevant, available information and, on that basis, comprehending information with respect to statutes in addition to the information encoded in statutory text. What is somewhat unique about statutes is the sometimes massive amount of information that we must draw on relative to that encoded in the communicative text. Indeed, in interpreting statutes it is common that we must actively seek out and use in inferences a large volume of information that appears nowhere in text. Further, we must often infer immense amounts of information relative to that which is textually encoded. Properly viewed, legal reasoning is the process of performing consciously and explicitly the mental operations, in large part inferences, that, in typical communicative settings, we perform unconsciously and automatically.

¹⁸ A full explanation of that interpretive process is reserved for a companion article.

PART I. THE JURISTIC BACKGROUND

The juristic lore holds that the source of the problems we encounter in interpreting statutes lies in an ambiguous ambiguity. Ambiguity is supposedly not only the *cause* but *the* cause of our interpretive problems. Since ambiguity is typically associated with words and phrases, some jump to the conclusion that the primary problem in the interpretation of statutes is purely lexical, relating to the meanings of words and phrases. That leads to the search for purely lexical solutions. However, scholars such as H. L. A. Hart, William Eskridge, Jr., and Richard Posner¹⁹ point out that sometimes the problem in understanding what a statute provides traces to the fact that the matter at issue was not contemplated when the statute to be interpreted was formulated. That is clearly not a problem attributable to lexical ambiguity. While both Professor Eskridge and Judge Posner have proposed an especially insightful "faithful agent" approach to these problems, neither has worked out in detail the full implications of his insights. As will become clear, that, in fact, is a significant part of what needs to be done.

Part II of this article explains in general terms the origin and nature of problems encountered in formulating statutory rule systems. Part III expands on the insights of the scholars identified in the preceding paragraph and demonstrates that many of the problems we encounter with respect to statutes are not at all lexical in nature. When that is so, the search for lexical solutions is completely beside the point. Part IV of this article demonstrates that there are, indeed, lexical problems associated with statutes, of which lexical ambiguity is the least. But, Part IV also demonstrates that if the lexical problems are not clearly understood, it is not possible to respond to them intelligently. For example, when faced with what they take to be a lexical problem, some justices and judges resort to dictionaries in what, as will become clear, is generally a vain

¹⁹ H. L. A. HART, *THE CONCEPT OF LAW* 124-36 (2d ed. 1994); RICHARD A. POSNER, *THE PROBLEMS OF JURISPRUDENCE* 269-73 (1990); William N. Eskridge, Jr., *Fetch Some Soupmeat*, 16 CARDOZO L. REV. 2209 (1995).

attempt to find the correct statutory meanings.²⁰ The first mistake these justices and judges, and many legal scholars for that matter, make is blithely to assume that dictionaries are sound sources for word meanings, although the fact dictionaries often differ should have given them pause. In contrast, linguistic scholars who specialize in lexical semantics appear to be unanimously agreed that dictionaries are not reliable sources for word meanings.²¹ Anyone who takes the time to

²⁰ There are several articles that discuss dictionaries as aids in statutory interpretation. See, e.g., Samuel A. Thumma & Jeffrey L. Kirchmeier, *The Lexicon Has Become a Fortress: The United States Supreme Court's Use of Dictionaries*, 47 BUFF. L. REV. 227 (1999) (an exhaustive article but for its failure to question the reliability and accuracy of dictionaries generally); A. Raymond Randolph, *Dictionaries, Plain Meaning, and Context in Statutory Interpretation*, 17 HARV. J.L. & PUB. POLY 71 (1994); Symposium, *Looking It Up: Dictionaries and Statutory Interpretation*, 107 HARV. L. REV. 1437 (1994).

²¹ "A definition expresses a hypothesis about the meaning of a particular word, and it is valid if it accounts correctly for the range of use of this particular word." ANNA WIERZBICKA, SEMANTICS: PRIMES AND UNIVERSALS 183 (1996). Unfortunately, "faith in dictionary definitions is based on an *illusion*." *Id.* at 250 (emphasis added). "[D]ictionary definitions don't offer anything like a faithful representation of a word's meaning" *Id.* "Often, they don't even try to provide a definition, but instead offer *more or less random lists of quasi-synonyms*" *Id.* at 240 (emphasis added). The problem, among others, is that the cost of producing truly reliable definitions would render any dictionary containing reliable definitions commercially unmarketable. *Id.* at 250. Equally important, it is not bloodless dictionary meanings that we need to know but the meaning of "words in the mind." It is "words in the mind," not words in dictionaries, that parties drafting statutes employ. However, "the mental lexicon is . . . a mammoth structure. The relationship between a book dictionary and the human mental lexicon may be somewhat like the link between a tourist pamphlet advertising a seaside resort and the resort itself." "[A] book dictionary gives us a *spuriously neat*, static and *incomplete* view of the mental lexicon. The differences between words in the mind and words in books are . . . *profound*." JEAN AITCHISON, WORDS IN THE MIND: AN INTRODUCTION TO THE MENTAL LEXICON 14 (1994) (emphasis added).

[B]ook dictionaries . . . serve a useful, though limited purpose, enabling us to check on the conventional spelling of a word and to find out its *approximate* meaning. Indeed, if we did expect a book dictionary to include the same information as the mental lexicon, then we would undoubtedly require a fleet of lorries [trucks] to carry its microstored edition around in.

Id. "[O]rdinary dictionaries are limited in scope in comparison with the mental lexicon. Their organization is oversimple, their content is fixed and outdated, and they contain only a relatively small amount of information about each item." *Id.* at

work out the meaning of a word along the principles outlined by Paul Ziff,²² Anna Wierzbicka,²³ and other experts in lexical semantics can confirm that fact for themselves.²⁴ So we are treated to the truly absurd spectacle of august justices and judges arguing over which unreliable dictionary and which unreliable dictionary definition should be deemed authoritative.²⁵

But there is a second and even greater mistake that these justices, judges, and legal scholars make. As Part IV makes clear, often the problem with words and phrases used in statutes is that the human lexicon does not at the time a statute is being formulated contain any word or phrase with a pre-existing meaning that is fully appropriate to the purposes of the statute at issue. When faced with that situation, the task of the court is not to pick any pre-existing meaning out of any dictionary, or even out of the aggregate existing mental lexicons of all English and non-English speakers, but rather, for the first time in human history to construct a meaning that is sensible in light of the purposes of the statute at issue and

15.

²² PAUL ZIFF, *SEMANTIC ANALYSIS* (1960).

²³ ANNA WIERZBICKA, *SEMANTICS: PRIMES AND UNIVERSALS* (1996); ANNA WIERZBICKA, *LEXICOGRAPHY AND CONCEPTUAL ANALYSIS* (1985).

²⁴ "[T]here are few things harder than constructing a good definition." ANNA WIERZBICKA, *SEMANTICS: PRIMES AND UNIVERSALS* 267 (1996). "[C]onstructing a definition which matches a word's entire range of use is a huge task requiring much more work than most are prepared to put in; and the temptation to give up after the first two minutes or the first two hours may well be overwhelming." *Id.* at 241. A good example of the process required to establish a sound definition is provided by Paul Ziff in his attempt to define "good" in English. *See* ZIFF, *supra* note 22. In doing so, Ziff examined 160 distinct uses of the word to insure that his definition adequately accounted for each. *See* ZIFF, *supra* note 22. For examples of definitions worked out systematically by a leading lexical semanticist, *see* ANNA WIERZBICKA, *LEXICOGRAPHY AND CONCEPTUAL ANALYSIS* (1985).

²⁵ Since lawyers and judges are unquestionably far more sophisticated about linguistic meaning than the typical layman, there is a tendency to believe we know more than we, in fact, do. It turns out that linguistic meaning is an extremely complex phenomenon. The author, through research in linguistics and reflection on that research, has identified more than twenty distinct linguistic phenomena that pass under the name "meaning." Jurists are neophytes compared with linguists, and it is time jurisprudence caught up with and incorporated the insights that linguistics has produced in over a century of scientific inquiry.

other purposes simultaneously operating in the law. When that is the task before the courts, resorting to dictionaries for ultimate answers is doubly absurd. Without knowing it, justices and judges are looking in dictionaries for the supposedly correct non-existent meaning.

On still other occasions, judges attempt to resolve lexical problems by resorting to supposedly grammatical or syntactical canons of construction.²⁶ Legal scholars have rightly in significant part discredited this procedure.²⁷ In addition, resorting to the supposedly grammatical or syntactical canons can also often be rejected on the grounds that contemporary linguistics and cognitive science have made it clear that in many instances the linguistic assumptions underlying those canons are positively false or hold only in special circumstances that cannot blindly be presumed to prevail in every legal setting.²⁸ For example, the canon that clear and unambiguous language requires no interpretation is scientifically wrong. Contemporary linguistics has established that, with the possible exception of some purely formal mathematical or logical statements,²⁹ all communicative language requires interpre-

²⁶ Justice Scalia is one who sings the praises of the grammatical, or textual, canons. See SCALIA, *supra* note 4, at 25-27. Wilson Huhn asserts that: "Textual canons . . . operate like rules of syntax in that they are used to infer the meaning of a rule from its textual structure." WILSON HUHN, *THE FIVE TYPES OF LEGAL ARGUMENT* 23 (2002).

²⁷ See, e.g., Edward L. Rubin, *Modern Statutes, Loose Canons, and the Limits of Practical Reason: A Response to Farber and Ross*, 45 VAND. L. REV. 579 (1992); KENT GREENAWALT, *STATUTORY INTERPRETATION: 20 QUESTIONS* 211 (1999).

²⁸ Neither have jurists provided unquestioned grounding for the substantive canons. As the article that is a companion to this article makes clear, with a little support from contemporary linguistics some of the substantive canons that are much maligned by textualists can be solidly grounded linguistically, epistemologically, and jurisprudentially (including constitutionally).

²⁹ Such statements are sometimes referred to as "eternal sentences." See ROBYN CARSTON, *THOUGHTS AND UTTERANCES* 29-32, 38-39, 65, 74-83 (2002). Jon Barwise and John Perry refer to such sentences as "nonefficient" in that they can convey only a single, timeless meaning:

One can imagine nonefficient sentences, sentences that make the same claim about the world no matter who says them and when. These are (roughly) what Quine calls *eternal sentences*. They contain no words like I, no demonstratives like THIS, no tense. One can imagine such sentenc-

tation, often at several levels.³⁰

In addition to the above, most justices and judges regularly resort to a somewhat vague legislative intent to find the meaning of words and phrases that legislators may have had in mind at the time of drafting.³¹ They also regularly resort

es, but with the possible exception of sentences from mathematics it is not easy to produce examples. . . . [T]he efficient unit of meaning (whether expressions or something else) is the norm.

JON BARWISE & JOHN PERRY, *SITUATIONS AND ATTITUDES* 6 (1983). Some argue that all sentences are efficient, i.e., require interpretation. See CARSTON, *supra*, at 30-42.

³⁰ See, e.g., DAN SPERBER & DEIRDRE WILSON, *RELEVANCE: COMMUNICATION & COGNITION* (2d ed. 1995); STEPHEN C. LEVINSON, *PRESUMPTIVE MEANINGS: THE THEORY OF GENERALIZED CONVERSATIONAL IMPLICATURE* (2000); CARSTON, *supra* note 29; A. J. SANFORD & S. C. GARROD, *UNDERSTANDING WRITTEN LANGUAGE: EXPLORATIONS OF COMPREHENSION BEYOND THE SENTENCE* (1981); Phillip N. Johnson-Laird, *The Mental Representation of the Meaning of Words*, 25 *COGNITION* 189 (1987). Even a sentence as simple as "the cat is on the mat" requires significant interpretation. If the sentence is uttered as an actual communication, it will mean, depending on exactly who says it, when, and where, something like: "The cat (that this speaker is referring to) is (at 10:46 a.m., on November 18, 2005, Central Time, U.S.A.) [elided—in (the speaker's dining room)] on the mat (that the speaker is referring to and that is on the speaker's dining room table)." A great deal of information, such as the specific cat, the actual time, and the actual place of reference (all the matter subscripted), that is "not said" would be understood not through decoding but through inferences from the totality of all available information.

Concerning the claim that words that are clear and unambiguous require no interpretation, there is the story related by Dan Sperber and Deirdre Wilson. In 1930, two friends were discussing the Soviet Union. One, an optimist, believed that Lenin and his successor, Stalin, had created the workers' paradise. The other, a pessimist, claimed it was all a fraud. The optimist realized that the only way to settle the dispute was to visit the Soviet Union. He said: "I will go to the Soviet Union, see how things are, and write you and tell you. Now, although I am an optimist, I realize things are not always as they seem. If the Soviet Union is a police state and I cannot report truthfully, I will write what they want me to write, but I will write in purple ink." The optimist left for his visit, and in due course a letter arrived. The letter was written in black ink. In it the optimist said that the Soviet Union was, indeed, the workers' paradise. All workers lived in castles and dined on caviar. His final sentences read: "Consumer goods are so plentiful and inexpensive you would not believe it. You can buy everything you want with only one exception. You cannot buy purple ink." In this context the words "you," "cannot," "buy," "purple," and "ink" are not in the least bit ambiguous. Nonetheless, in the context above, "you cannot buy purple ink" means: "The Soviet proclamations of the workers' paradise are lies." See SPERBER & WILSON, *supra*, at 170.

³¹ The author has not yet found an adequate explanation of the concept of

to statutory purposes as an aid in understanding both the meanings of words and phrases and entire statutory provisions.³² Both these procedures are cognitively, linguistically, epistemologically, and jurisprudentially, including constitutionally, proper and sound—as far as they can take us. The problem is that the traditional resorts to legislative intent and statutory purposes have their limitations.³³ Again, drawing on the insights of scholars such as Hart, Eskridge, and Posner, if the matter at issue was not even remotely contemplated by the parties formulating the statute, there will certainly be no express legislative intentions regarding it in any of the traditional sources. Likewise, the statutory purposes ascertainable by traditional means, such as text and legislative history, will also fail to shed light on the matter. Traditional jurisprudence has no way of ascertaining the intentions of legislators in such circumstances and, inevitably, comes up short.³⁴

There is, in addition, an array of interpretive alternatives, none of which fully addresses the problems encountered in statutory interpretation, has achieved general acceptance, or clearly accords with the constitutional provision for legislative supremacy.³⁵

legislative intent. Sometimes the term appears to be limited to the intention to supply a specific meaning for a word or phrase. On other occasions it seems to include the intention to pursue specific purposes. In the instant discussion, nothing turns on sorting matters out.

³² This is a time honored approach evidenced in Edmund Plowden's theory as set forth in his 1574 report of *Eyston v. Studd*, (1574) 75 Eng. Rep 688 (K.B.), and the commentary in the report of the Exchequer Chamber's opinion in *Heydon's Case*, (1584) 76 Eng. Rep. 637, 638 (Exch.).

³³ See WILLIAM N. ESKRIDGE, JR., DYNAMIC STATUTORY INTERPRETATION 13-47 (1994).

³⁴ There are, indeed, intentions, a relatively complex set of intentions, that we can, indeed, must, attribute to the legislature in these situations. That set is introduced briefly in the Conclusion to this article and examined in detail in the companion to this article. See J. Gordon Christy, *Statutory Interpretation: The Gordian Knot Untied* (unpublished manuscript, on file with the author). This article is intended to loosen some of the strands of that knot.

³⁵ For a masterful and balanced critique of interpretive theories, see ESKRIDGE, *supra* note 33. Professor Eskridge clearly proves that statutory interpretation is dynamic, but he does not explain in full why it must be dynamic and

Without understanding exactly what it is they are confronting, and reflecting on the chaotic state of jurisprudential interpretive theory, some scholars conclude that nothing constrains interpretive outcomes when statutes are at issue.³⁶ And, as noted above, at least one obviously very intelligent and deep thinking scholar has concluded that there is no right way to interpret statutes. Nonetheless, most jurists, especially practicing lawyers, adopt a sort of "Humean" attitude³⁷ and manage to work within the system, finding that despite the chaos in interpretive theory, there is sufficient predictability in the way judges decide cases so that competent lawyers can give generally reliable advice. Judicial behavior is sufficiently consistent that outcomes are not all random—and neither are they all determined by the political proclivities of judges. Wilson Huhn has ably explained why that is so.³⁸ Specifically, he has shown that there are identifiable types of arguments that courts consider in interpreting statutes, and sometimes all or most of the arguments lead to the same result. Accordingly, even if traditional legal reasoning does not *compel* specific outcomes, it, nonetheless, very often has the practical effect of producing quite predictable and specific outcomes. Nevertheless, individual litigants are at mercy of a multitude of inter-

how to circumscribe the dynamism in a way that fully accords with the constitutional principle of legislative supremacy.

³⁶ See, e.g., Joseph William. Singer, *The Player and the Cards: Nihilism and Legal Theory*, 94 YALE L.J. 1 (1984); Gary Peller, *The Metaphysics of American Law*, 73 CAL. L. REV. 1151, 1152 (1985); David Kairys, *Law and Politics*, 52 GEO. WASH. L. REV. 243, 244 (1984).

³⁷ David Hume, one of the giants in modern philosophy, argued quite persuasively that, inter alia, one cannot empirically prove that physical objects persist when unobserved or discover any necessary connection between causes and effects. DAVID HUME, A TREATISE OF HUMAN NATURE pt. IV (L. A. Selby-Bigge ed., 2d ed. 1978) (1739-40). That seemed to him to leave open the possibility that the world could literally dissolve into chaos. *Id.* He admitted, however, that as soon as he left philosophizing behind and entered the company of good friends he promptly forgot all about the problem. *Id.* "I dine, I play a game of back-gammon, I converse with my friends; and when after three or four hour's amusement, I wou'd return to these speculations, that I cannot find in my heart to enter into them any further." *Id.* at 269.

³⁸ HUHNS, *supra* note 26, at 83-89.

pretive theories. The very multiplicity of theories is itself one of the primary causes of indeterminacy in the law. Since different judges adhere to different theories, the random assignment of judges, even when panels of judges are convened, on many occasions unquestionably gives rise to randomly different outcomes. There is, thus, a regrettable, residual randomness in justice delivered. Justice is also more expensive than it would be if we had a uniform theory of interpretation that increased the percentage of predictable results. As things stand, many litigants have incentive to play the costly interpretive theory lottery.

PART II. THE REAL PROBLEMS IN STATUTORY INTERPRETATION—AN OVERVIEW

Despite the juristic lore, ambiguity, in any of its aspects, is only most rarely the cause of the problems we face in interpreting statutes. It is the practical problems we encounter in formulating statutory rule systems that virtually always cause our problems.³⁹ There are, as noted above, three primary practical problems that together give rise to at least twenty-

³⁹ This became clear to the author as a result of his practice experiences. The author entered the practice of law in 1975, just as the Employee Retirement Income Security Act ("ERISA") was taking effect. The provisions of what was originally ERISA are currently largely found at 29 U.S.C. §§ 1104-1144, 1301-1461 (2000) and 26 U.S.C. §§ 401-415, 4974-4975 (2000). The author spent over twenty years in practice watching ERISA's rule system grow through regulations, administrative agency rulings and opinion and advisory letters, and case law. *See infra* Part III. In 1993 the author assumed responsibility for advising a major corporation of its responsibilities under the then recently enacted Family and Medical Leave Act of 1993, 29 U.S.C. §§ 2601-2654 (West 1999). The same phenomenon repeated itself over the next ten years, albeit on a significantly smaller scale. It then became clear to the author that all statutes undergo the same evolution, an evolution that is not all that different from the evolution of the common law. The only substantial difference is that the legislature attempts from the beginning to paint more broadly than do the courts. But, in spite of the running start, statutes almost immediately raise problems identical to those encountered in the common law.

Lawyers, in fact, run into the identical problems in the drafting of contracts in connection with complex transactions and in drafting pension plans and stock option plans. Addressing in advance all the problems that might surface is simply impossible.

nine distinct sub-problems. Conceptual, lexical ambiguity is only one of those problems.⁴⁰ The problems that we indiscriminately call "ambiguity" are virtually always secondary problems that flow from the primary, practical problems. Statutory interpretation is, in significant part, addressing these practical problems.

To expand on what was said above, the first practical problem is that the subject matters underlying statutes are too complex, our initial understanding of those subject matters often too limited, and the competing purposes of the law almost always too complex to permit the *ex ante* formulation of fully refined statutory purposes. Rather, all statutory rule systems inevitably emerge with underlying purposes that are in various degrees unrefined. Occasionally, the underlying purposes are in significant respects inchoate. The second practical problem is that the incomplete initial understanding of the underlying subject matters, the less than fully refined purposes, the potentially infinite detail that needs to be addressed by all statutes that address complex subject matters, and the interplay of every statute's purposes with other purposes the law is simultaneously pursuing (but that are largely beyond the horizon of awareness of the parties formulating any given statute as they focus on the problems at hand) render it impossible *ex ante* to formulate anything even approaching a complete statutory rule scheme. Even if statutory purposes are relatively completely refined, the remaining obstacles mean that the statutory rule systems that are designed to achieve them inevitably emerge incomplete in greater and lesser degrees. In fact, any relatively complex statutory rule system will, in some respects, forever remain incomplete.

The third practical problem is that the issues we attempt to address through statutes are so varied and complex that it

⁴⁰ There is a thirtieth problem we encounter in the interpretation of statutes. That is syntactical ambiguity. For an excellent discussion of syntactical ambiguity see SOLAN, *supra* note 16. This article does not discuss syntactical ambiguity because such ambiguity can always be eliminated by careful drafting. The problems with which this article is concerned are those inherent to the formulation of statutes and those that remain after even great care is taken in drafting.

is *ex ante* impossible to ascertain the meanings for words and phrases that would be perfectly appropriate to the purposes of statutory rule systems and other purposes that the law is simultaneously pursuing. Accordingly, statutes inevitably emerge containing words and phrases with common meanings or even specially drafted definitions that are not in all ways appropriate to the purposes underlying that statute and other purposes the law is simultaneously pursuing.

The unfortunate fact is that only an omniscient and fully prescient mind could *ex ante* fully understand the kinds of complex subject matters that tend to underlie statutes, fully refine a statute's purposes in light of a complex underlying subject matter, take into account all the competing purposes of the law, and formulate a complete statutory rule system incorporating definitions for words and phrases that are fully appropriate to all those purposes. And if the mind of such a being worked at the speed of a typical human mind, it would still be a lengthy process—one that on some occasions might not be possible to complete in a politically acceptable amount of time. The consequence is that when we encounter situations that the parties drafting a statute overlooked or failed to anticipate and in these situations apply statutory rule systems literally, that is, give effect to their presumptive semantic content, the results are not only often unintended, accidental, random, and, accordingly, arbitrary, but also, often, in varying degrees, irrational, unfair, or unjust.⁴¹ The latter outcomes conflict with either (or both) the purposes underlying the statute or other purposes simultaneously operating in the law. We label such conflicts "ambiguity" without specifying the type of ambiguity at issue. In fact the ambiguity is fundamentally an ambiguity of purpose, the primary type of ambiguity we encounter with respect to statutes. These conflicts arise, or slip through, because the parties formulating a statute are not aware they potentially exist and, therefore, both cannot and do not address them.

⁴¹ That does not mean that they are *always* unacceptable, only that they are highly likely to be.

Part III of this article identifies four types of ways in which the purposes underlying statutory rule systems can emerge unrefined and twelve types of ways in which statutory rule systems inevitably emerge incomplete.⁴² Each of these types can occur many times with respect to a single statute. To put the matter in terms familiar in linguistics, there are in total at least sixteen types of ways that statutory purposes and rule systems emerge unrefined and incomplete, and each of those types may surface in numerous tokens. The result is that the number of actual instances of incompleteness can run into the hundreds and thousands.⁴³

Part IV identifies thirteen ways in which statutory rule systems inevitably begin life with key words and phrases that are partially inappropriate to the purposes underlying the statutes and other purposes simultaneously operative in the law. As is the case with statutory purposes and rule systems, in the absence of an omniscient and fully prescient mind, the partial impropriety of the meanings of words and phrases is not practically recognizable, or at least not fully recognizable, *ex ante* and cannot be rectified before a statute is enacted. Again, each type of partial impropriety may surface numerous times, giving rise to tokens that in number are many multiples of twelve, occasionally hundreds or thousands of tokens.⁴⁴

⁴² The author makes no claim that any of the lists set forth in this article is exhaustive.

⁴³ If this assertion, or at least the "thousands" part, strikes the reader as exaggerated, the author suggests that the reader turn to the discussion of the Employee Retirement Income Security Act of 1974 (the provisions of what was originally ERISA are currently largely found at 29 U.S.C. §§ 1104-1144, 1301-1461 (2000) and 26 U.S.C. §§ 401-415, 4974-4975 (2000)). See *infra* Part III. And after reading that discussion, count the number of United States Code Annotated annotations following each of the provisions of the Act.

⁴⁴ The terms "management" and "administrative" under the Fair Labor Standards Act regulations (29 C.F.R. §§ 541.1, 541.2, 541.102-541.115, 541.200-541.210 (2005)) were so vague that the courts treated their application as one for the fact finder. In literally thousands of instances one could not tell whether specific jobs fell inside or outside the categories. Had the courts bitten the bullet and undertaken to eliminate the vagueness, the resulting definitions would have been exceedingly complex, indeed.

The foregoing practical problems give rise to two linguistic problems, only one of which is lexical. Given the complexity of the subject matters and purposes underlying a statutory rule scheme and other purposes simultaneously operative in the law (which, to repeat, are in large part beyond the horizon of awareness of the parties formulating the statutory rule system at issue), parties formulating statutory rule systems cannot uncover and anticipate all necessary and/or appropriate rules (including exceptions to rules and resolutions of conflicts between rules and purposes) and will inevitably fail to include some, often many. They will also incorporate words and phrases with common meanings and even express statutory definitions that are not fully appropriate to the purposes underlying the rule scheme and other purposes simultaneously operative in the law. Even if they become aware of the need to formulate special statutory definitions, seldom will those definitions be adequate in all respects. The linguistic consequence, rather obviously, is that the texts of statutes do not contain all the sentences they would if they were complete—sentences that would set forth additional substantive and remedial rules. And, to reiterate, the sentences that do appear in the text will contain words and phrases with apparent meanings, i.e., their common meanings or specially drafted definitions, that are not fully appropriate to the statutes in which these words and phrases appear.

PART III. THE INHERENT INCOMPLETENESS OF STATUTORY RULE SCHEMES

Cognitive science, with assistance from the philosophy of mind and general principles of epistemology, has, over the last twenty or so years, made clear the central role that our knowledge of intentions and purposes plays in the comprehension of all human behavior, including linguistic behavior.⁴⁵ Building

⁴⁵ See, e.g., SPERBER & WILSON, *supra* note 30; CARSTON, *supra* note 29; SIMON BARON-COHEN, *MINDBLINDNESS: AN ESSAY ON AUTISM AND THEORY OF MIND* (1995); ROBERT AUDI, *ACTION, INTENTION, AND REASON* (1993); DANIEL C. DENNETT, *THE INTENTIONAL STANCE* (1987); CHARLES TAYLOR, *THE EXPLANATION*

on these insights, the linguistic discipline of pragmatics⁴⁶ has focused intensively on the central role that inferential processes that make use of communicators' intentions and purposes play in language, the creation linguistic meaning(s), and the formulation and comprehension of linguistic communications.⁴⁷ Pragmatics has made it clear that, with the possible exception of certain statements in pure logic or mathematics, understanding any linguistic communication requires taking the communicator's intentions and purposes into account.⁴⁸ As noted above, contrary to the judicially declared canon that proclaims that unambiguous language precludes the need for interpretation, scientific linguistics has made it clear that plain and unambiguous language never precludes the need for interpretation. Words, phrases, and sentences, spoken or written, are always only evidence, along with other evidence, especially of communicators' intentions and purposes, of what a speaker or writer means, in the sense of "intends the recipient of his or her linguistic communication to comprehend."⁴⁹

The fact is that the human cognitive system is as much wed to intentions and purposes in the explanation and comprehension of human behavior, including linguistic behavior, as it is to mathematics in the explanation and comprehension of physical phenomena. Intentions and purposes are cognitive, explanatory vehicles that operate in our understanding and explanation of human behavior in a way that is analogous to the way forces operate in our understanding of physical phe-

OF BEHAVIOUR (1964). See also Noam Chomsky's devastating 1957 attack on behaviorism. Noam Chomsky, *A Review of Skinner's Verbal Behaviour*, 35 LANGUAGE 26 (1959). For an introduction to the interrelations between linguistics, philosophy, and cognitive science see, for example, HOWARD GARDNER, *THE MIND'S NEW SCIENCE: A HISTORY OF THE COGNITIVE REVOLUTION* (1985) (a cognitive scientist's perspective); OWEN J. FLANAGAN, JR., *THE SCIENCE OF THE MIND* (1984) (a philosopher's perspective).

⁴⁶ The contemporary linguistic discipline of pragmatics, not to be confused with the philosophical school of pragmatism, derives originally from the writings of the linguistic philosopher H. Paul Grice.

⁴⁷ See *supra* note 30.

⁴⁸ See *supra* note 30.

⁴⁹ This is, indeed, the central of the many meanings of "meaning."

nomena.⁵⁰ It has become clear that, whether or not intentions, purposes, causes, numbers, and mathematical relations exist as phenomena distinct and apart from the human cognitive faculty, they are the tools with which we must operate in their distinctive spheres. It would make as much sense to demand that we understand human behavior, including linguistic behavior (which includes statutes), without the explanatory vehicles of intentions and purposes as it would to demand that we understand the world of physics without resort to mathematics and mathematical relations.⁵¹

All that said, to repeat what was noted above, one of the three primary, practical, and therefore communicative, problems relating to statutes flows from the fact that it is *ex ante* virtually impossible to fully formulate the purposes of a statutory rule scheme with an even modestly complex underlying

⁵⁰ For an explanation of intentions (which have related purposes) as causes see, for example, JOHN SEARLE, *INTENTIONALITY* 112-40 (1983). Steven Pinker explains the role of intentions and purposes in terms of "feedback loops." STEVEN PINKER, *THE BLANK SLATE: THE MODERN DENIAL OF HUMAN NATURE* 32 (2002). See also AUDI, *supra* note 45, at 110-144, for a discussion of the nomological, even if not causal, nature of intentions.

⁵¹ Some argue that we cannot sensibly ascribe intentions and purposes to the legislature. There are so many minds. How can they all have the identical intentions or be set on the identical purposes? One answer is that collective action in large groups is never the result of absolutely identical states of mind of all group members. Of necessity, some small group out of the whole, often a single individual, has to be designated to provide the intentions and purposes whereby the group acts. So the legislative intentions and purposes with which we are concerned of necessity must, in part, be those of the key actors in crafting and moving legislation forward. The key actors propose and the remainder disposes, whether the remainder understands or even cares about, what has been proposed. Their knowledge and concern is not constitutionally required. But there is an even more fundamental reason we must ascribe intentions to the legislature. As noted above in the text, intentions and any accompanying purposes are the only vehicles the human cognitive system provides us for understanding and explaining human behavior. We must use, and cannot operate without, that which nature has given us, even if what nature has given us is not, in the setting of the law, in every respect ideal. Whether it is conceptually neat to do so, we have no choice but to analyze legislative actions in terms of intentions and purposes. We find ourselves in the same bind with mathematics. We might prefer more powerful tools in addressing relations between physical phenomena, especially in particle physics and cosmology, but our cognitive faculties provide us only with mathematics.

subject matter. No wonder we are sometimes stymied in attempting to understand in full the purposes of a statute. The parties who formulated it did not themselves fully understand them!

But even if we were *ex ante* able to fully understand and formulate the purposes of a statute with all proper refinements, we still could not *ex ante* formulate in full a corresponding rule system to address even a minimally complex underlying subject matter. The problem is compounded to the extent the purposes are unrefined. Reality and the cross currents of the purposes of the law are too varied. This becomes clear by considering what would be required *ex ante* to produce a complete statutory rule system. The significance of what follows will be obvious to anyone who has drafted a document such as a complex real estate construction contract or corporate merger agreement. The number of issues that could be addressed is endless. In that light, the legislature could produce a complete statutory rule system only if it were capable of, *inter alia*:

- (1) thoroughly mastering the underlying subject matter,
- (2) thoroughly refining the purposes to be achieved through the statute,
- (3) (a) identifying every rule that is necessary or that would be beneficial in furthering or achieving the statute's purposes (or ends) in all possible situations, without regard to the rarity, remoteness, or uniqueness of any of such situations, (b) identifying every rule that would interfere or conflict with the furthering or achieving of the statute's purposes, including, in particular, all such rules found in the United States Code, state laws, or derivable from our traditions and customs, and (c) identifying, with respect to the set of rules in (3)(b), each rule that, in the legislature's view, should be subservient to and every rule and purpose that should override the ends of the statute at issue, and
- (4) taking all the above into account, (a) except as provided in (b), immediately following, incorporating into the statute rules, both substantive and remedial, relating to all possible clusters of factual occurrences, without regard to the rarity,

remoteness, or uniqueness of such clusters, necessary or beneficial in furthering or achieving the statute's ends, (b) expressly excluding from those rules "merely" beneficial rules that the legislature deemed to be (i) too harsh in their consequences given their small or virtually inconsequential contribution to the ends sought, (ii) too inconsequential or peripheral to be worth devoting judicial resources to, or (iii) to conflict with more important, overriding ends, and the like, (c) including, if not covered above, all subsidiary rules required in applying the primary rules,⁵² (d) providing exceptions for all rules that, if literally applied, would on rare occasions inhibit the furthering or achieving of the statute's ends, with the exception of those necessary to protect ends deemed to override those of the statute, which rules would, of course, be included, (e) including all exceptions to the general rules necessary to prevent what the legislature deemed to be irrational, unfair, and unjust results in light of unique circumstances, (f) expressly resolving all internal conflicts between statutory provisions and all conflicts between the statute and, if not addressed above, all other statutes and the Constitution, (g) expressly sanctioning or denying effect to all results flowing from the simultaneous or successive application of separate rules, wherein results can be achieved that could not be achieved through application of the rules singly,⁵³ and (h) making it clear exactly who is entitled to seek the remedies provided by the statute, i.e., whether the statute permits private causes of action.⁵⁴ Further, to render a statute truly

⁵² See *infra* pp. 85-86 and 90-92 and *infra* note 105.

⁵³ For example, through *seriatim* transactions that each comply with the requirements of the "tax free" reorganization provisions of the Internal Revenue Code, parties can effect the distribution of earnings and profits taxable as dividends in the form of capital gains. See BORIS I. BITTKER & JAMES S. EUSTICE, *FEDERAL INCOME TAXATION OF CORPORATION AND SHAREHOLDERS*, ¶¶ 1.05[2][d]. 12.61[3] (6th ed., Warren, Gorman & Lamont 1994); Marvin A. Chirelstein & Benjamin B. Lopata, *Recent Developments in the Step-Transaction Doctrine*, 60 TAXES: THE TAX MAGAZINE 970 (1982); Joseph Isenbergh, *Musings on Form and Substance in Taxation*, 49 U. CHI. L. REV. 859 (1982).

⁵⁴ See, for example, the principles discussed in *Alexander v. Sandoval*, 532 U.S. 275 (2001); *Jackson v. Birmingham Bd. of Educ.*, 309 F.3d 1333 (11th Cir.

complete, the legislature would have to assign all words and phrases used in the statute meanings perfectly appropriate to the statute's ends and all other purposes the law is simultaneously pursuing, neither covering facts that should not be covered nor excluding facts that should be covered. Thus, either the historical meanings of words and phrases would have to be perfectly suited to the needs of the statute or the legislature would have to recognize any inadequacies in those meanings and address them with special, nonce, statutory definitions. In the case of words and phrases that have more than one meaning, including prototypical and non-typical meanings, if one meaning perfectly suited the needs of the statute, it would be necessary to specify which. These issues relating to words and phrases are addressed in Part IV.

Although to an individual who has never undertaken anything comparable, the tasks above might, in the abstract, seem difficult but hardly overwhelming, experience proves conclusively that they are overwhelming. Statutes always and inevitably emerge from the drafting and enactment processes varying in incompleteness from the extremes of the "common law" statute to what the Supreme Court, confusing complexity with completeness, has called the "comprehensive and reticulated" statute. Assuming that such extremes are accurate and meaningful categories, given the massive incompleteness of the paradigmigmatic so-called "comprehensive and reticulated" statute, discussed below, a statute that emerged lacking literally thousands of pages of rules and no telling how many specific rules, there are in between what may be called "typically incomplete" statutes. Obviously, the degree to which a statute is substantively or remedially incomplete is a function of the difference between the rule system that would truly be complete and the rule system actually provided by the text of the statute as well as any differences between the historical meanings of the words and phrases used in the statute or their statutory definitions and the meanings actually required in light of the statute's purposes.

Again, it is crucial to recognize that the incompleteness of statutes is almost always not intended by the legislature.⁵⁵ Consequently, virtually all of the omissions relating to any given statute are random and accidental. When that is the case, giving effect to the literal content of the statute cannot but yield purely arbitrary and accidental results.

The following illustrates concretely, and hopefully with a high degree of thoroughness, the endemic nature of the problems in the cases of "common law" statutes, so-called "comprehensive and reticulated statutes," and typical statutes.

A. "Common Law" Statutes

The paradigmatic "common law" statute is, of course, the Sherman Act,⁵⁶ a legislative attempt to deal, at the very least, with socially harmful, anticompetitive business conduct.⁵⁷ The rule scheme envisioned by the Sherman Act was

⁵⁵ It is obvious that the legislature occasionally ducks politically sensitive issues and leaves resolution to the courts.

⁵⁶ 15 U.S.C. §§ 1-11(2000).

⁵⁷ Some contemporary scholars claim that the Sherman Act was not concerned with the dangers posed to democracy by massive aggregations of wealth (that, in spite of the fact that with enough wealth one can buy the repeal of the Sherman Act, which may well have effectively occurred during the Reagan Administration), that unrestrained competition is somehow "perfect" (a claim that would probably have convulsed Adam Smith in paroxysms of laughter), and that efficiency and efficiency alone was the concern of Congress in passing the Sherman Act. The author finds those claims incredible. The agricultural interests of the United States in 1890 clearly saw the dangers of aggregations of wealth. Not that they were disinterested observers. Their once dominant powers were being severely eroded. To think that they were not concerned about political power when they were both losing it and exercising it to secure passage of the Act is absurd. Further, to legislators raised on classical history, the lessons of the effects on Roman society of the destruction of the independent, citizen farmer (the Roman equivalent of the small, independent businessman) that formed the backbone of the Roman Republic would have been clearly before their minds. The author, accordingly, believes that concern with excessive financial and political power was almost certainly a concern underlying the Sherman Act. The author believes that Robert Bork and others who focus solely on efficiency may be unwittingly willing to trade American freedom for a mess of pottage in the form of a little short-term economic efficiency. It may well be that we cannot have both maximum economic efficiency and political freedom, but rather that political freedom is not free and part of its price is an occasional and *moderate* reduction in efficiency.

initially highly incomplete for the most obvious of reasons. Congress did not and could not, under the circumstances, begin to fully understand the nature of the problems it faced, was not certain about all the problems it wanted to address, and, accordingly, could not give the courts more than general directions to deal with certain aspects of the untoward fall-out of the industrial, commercial, and financial developments of the Second Industrial Revolution in America. The purposes of the Act were, therefore, in significant ways inchoate.

Congress knew that there were potential problems. The post-Civil War technological advances in industrial and agricultural production, combined with the rail connection of the East and West Coasts, the resulting opening up of the Great Plains, and significant advances in transoceanic transportation, made possible mass markets that, in turn, gave rise to potential economies of scale in financing, production, marketing, and distribution not previously possible. Economies of scale produced economic blessings in the form of reduced costs. To take advantage of those economies of scale required aggregations of capital on an unprecedented scale. The new capital markets, together with the legislatively created limited liability business corporation, the creative use of that antique device, the trust, to effect combinations of previously independent production units, and finally the general business corporation acts with provisions for merger and consolidation effectively at will,⁵⁸ resulted in financial and market power, labor market effects, and other consequences previously unknown.

Many recognized that these developments were not all necessarily for the good. Massive financial power meant mas-

⁵⁸ New Jersey enacted the first "modern" general business corporation act in 1875. Adolf A. Berle, Jr., *Historical Inheritance of American Corporations*, in *Social Meaning of Legal Concepts*, 3 THE POWERS AND DUTIES OF CORPORATE MANAGEMENT 189 (1950). Under that Act the restrictions on the business activities that could be conducted through a single corporation and on the ability to merge enterprises and bring disparate businesses under common control were effectively removed. Until that Act the most effective way to combine separate corporate enterprises, whose purposes were limited under special charters, under unified control was through the common law trust. After the New Jersey Act came on the scene, the trust was rendered relatively obsolete.

sive political power. There was no guarantee that either would be exercised for the general welfare. It was also clear that smaller, less efficient producers were being driven out of the market. Individuals who had been educated in the Classics and knew that the demise of the independent Roman farming class had doomed the Republic could not but worry that the reduction in the number of independent producers would be inimical to American democracy.⁵⁹

In addition to all the foregoing, no one fully understood the new economy that was developing. In fact, the analytical tools for understanding it were only in their infancy. Alfred Marshall's *Principles of Economics*, the first mathematically oriented and comprehensive treatment of what is now known as micro-economics, appeared in 1890, the year of the Sherman Act.⁶⁰ And until the discipline of micro-economics advanced well beyond Marshall's initial insights, it would prove impossible to fully understand the economic issues underlying the act. The common law had developed rules against unreasonable restraints on trade, and the Seventeenth Century Parliament had also outlawed all monopolies except those it specifically authorized.⁶¹ But the common law concepts of "unreasonable restraint" and "monopoly" were neither clearly sufficient nor clearly insufficient to deal with the new phenomena, and, if insufficient, it was not clear exactly how.

⁵⁹ Justice Peckham voiced a similar concern in *United States v. Trans-Missouri Freight Association*, 166 U.S. 290, 323-24 (1897). See also *Richardson v. Buhl*, 77 Mich. 632, 43 N.W. 1102, 1110 (1889) (Chief Justice Sherwood observing: "[I]t is doubtful if free government can long exist in a country where such enormous amounts of money are allowed to be accumulated in the vaults of corporations."); 21 Cong. Rec. 3, 2598 (1890) (remarks of Senator George regarding the potential elimination of all small capitalists). It was also obvious to labor that capital was gaining immense advantages over labor. The corporation was simply a union with another name and some special legal advantages. Capital could freely form unions, called corporations, but labor could not. Capitalists could join in single enterprises and the combine their single enterprises under trusts and holding companies, and finally into single enterprises through merger. The capitalists recognized their immense advantages over labor and effectively fought, with the assistance of both Congress and the courts, to maintain them until the 1930s.

⁶⁰ ALFRED MARSHALL, *PRINCIPLES ECONOMICS* (1890).

⁶¹ Statute of Monopolies (1623).

After a short period of intense confusion,⁶² the judicial response was to treat many kinds of business conduct as per se violative of the Act. More recently, that is, around eighty years after the Act became law, as the discipline of micro-economics has advanced and judges have become better educated in economic matters, much conduct that was proscribed in the era of the broad per se rules has been found to pass muster under a much more flexible and sophisticated "rule of reason." Today it is even recognized, if only tacitly, that in the rarest circumstances reasonable price fixing is unavoidable, in fact, preferable to free market competition, and, therefore, lawful.⁶³

The resulting economic, social, and political phenomena that for the first time emerged after the Civil War ultimately had to be recognized, understood, and addressed over a period of seventy to eighty years through legislation such as the Sherman Act, the Interstate Commerce Act,⁶⁴ the Clayton Act,⁶⁵ the Federal Trade Commission Act,⁶⁶ Robinson-Patman Price Discrimination Act,⁶⁷ state public utilities acts,

⁶² The United States Supreme Court, in its first decision under the Act, held that the Sugar Trust, which had acquired control of ninety-percent of the sugar manufacturing in the United States, although a virtual monopoly, was not covered by the Act because the Act addressed attempts to monopolize commerce, and manufacturing is distinct from commerce! *United States v. E.C. Knight Co.*, 156 U.S. 1 (1895). In *United States v. Trans-Missouri Freight Association*, 166 U.S. 290 (1897), the court held that the Sherman Act renders illegal *all* restraints on trade. Of course, anyone who contracts with another in part restrains his ability to contract with others and, thereby, somewhat restrains trade. That would render all contracts illegal restraints on trade. The court rather quickly backed off and held that the Act proscribes only "unreasonable restraints" on trade. *United States v. Addyston Pipe & Steel Co.*, 85 Fed. 271 (1898) (the opinion of Judge Taft, later Chief Justice Taft, was effectively adopted by the United States Supreme Court).

⁶³ At least that is the way the author reads *Broadcast Music, Inc. v. Columbia Broadcasting System, Inc.*, 441 U.S. 1 (1979).

⁶⁴ 24 Stat. 379 (ch. 104) (Feb. 4, 1887) (as amended on over one hundred occasions).

⁶⁵ 15 U.S.C. §§ 12-17 (2000).

⁶⁶ 15 U.S.C. §§ 41-59 (2000).

⁶⁷ 15 U.S.C. §§ 13-13b, 21a (2000).

the Securities Act of 1933,⁶⁸ the Securities Exchange Act of 1934,⁶⁹ the Public Utilities Holding Company Act,⁷⁰ the National Labor Relations Act,⁷¹ the Labor-Management Relations Act,⁷² the Fair Labor Standards Act,⁷³ a host of other laws, and thousands of cases. Knowledge and understanding increased with time and so did the complexity, completeness, and, occasionally, the sophistication and effectiveness of the rule schemes provided by statutes.

At the moment of its enactment, the Sherman Act, accordingly, could not practically be more than a command to the federal courts to develop a federal common law of permissible and impermissible competitive and anti-competitive business behaviors as the courts, and the academy for that matter, *might in time come to understand them*. The statute, in many respects, could not be more than that because neither Congress nor the academy yet had the economic and financial understanding necessary for formulating anything but the barest directions for the construction of a rule system governing competitive and anti-competitive behaviors in the setting of the Second Industrial Revolution and its accompanying financial, marketing, and distribution revolutions. The courts had to create by far the largest portion of the rule system that now comprises the law of the Sherman Act, and the process has not reached completion today.

B. "Comprehensive and Reticulated" Statutes

Virtually everyone recognizes that the relatively rare "common law" statute is highly incomplete. What does not seem to be generally recognized is that all statutes addressing complex matters are incomplete, often highly incomplete. This fact can be readily illustrated in the case of the Employee

⁶⁸ 15 U.S.C. §§ 77a-77z, 77aa (2000).

⁶⁹ 15 U.S.C. §§ 77b-78ll (2000).

⁷⁰ 15 U.S.C. §§ 79-79z, 79z1-79z6 (2000).

⁷¹ 29 U.S.C. §§ 151-169 (2000).

⁷² 29 U.S.C. §§ 141-144, 151-167, 171-187 (2000).

⁷³ 29 U.S.C. §§ 201-219 (2000).

Retirement Income Security Act of 1974 ("ERISA").⁷⁴ The Supreme Court, confusing complexity with completeness, has on several occasions erroneously called ERISA a "comprehensive and reticulated" statute.⁷⁵ But, as the following discussion explains, ERISA emerged lacking literally thousands of pages of necessary and appropriate rules and no telling how many specific rules per "missing" page. Given the complexity of the subject matters ERISA addresses, the rule system embodied in the statutory text per se is neither comprehensive nor highly reticulated, although it is complex. ERISA's rule scheme, as textually stated, is, in fact, highly skeletal.⁷⁶

For example, many rules necessary for ensuring that one of ERISA's paramount purposes, i.e., protecting and providing for the sound management of financial resources set aside to provide benefits to pension⁷⁷ and welfare plan⁷⁸ participants, are missing. ERISA refers to such financial resources as "plan assets."⁷⁹ However, the term, "plan assets," given the historical meaning of the term "assets," was not fully appropriate to ERISA's purposes. Twelve years after ERISA's enactment, the

⁷⁴ The provisions of what was originally ERISA are currently largely found at 29 U.S.C. §§ 1104-1144, 1301-1461 and 26 U.S.C. §§ 401-415, 4974-4975 (2000).

⁷⁵ *Mertens v. Hewett & Assocs.*, 508 U.S. 248, 251 (1993); *Massachusetts v. Morash*, 490 U.S. 107, 113 (1989); *Alessi v. Raybestos-Manhattan, Inc.*, 451 U.S. 504, 510 (1981); *Nachman v. Pension Benefit Guar. Corp.*, 446 U.S. 359, 361 (1980).

⁷⁶ To truly understand the subject matters that ERISA addresses, especially employer sponsored pension and health plans, requires a degree in actuarial science, mathematics, or physics. The best pension/health plan consultants the employer worked with in private practice had at least a master's degree in mathematics, and one held a doctorate in mathematics while another held a doctorate in physics. All of these individuals admitted that the mathematical relations they worked with in plan design are so complex that they could not intuit answers but, rather, actually had to "crunch" the numbers. That being the case, Congress was certainly not going to provide all the rules necessary for governing pension and health plans.

⁷⁷ For purposes of ERISA, a pension plan is any plan, whether traditional defined benefit plan or not, that defers compensation to the termination of employment or retirement. 29 U.S.C. § 1002(2)(A) (2000).

⁷⁸ Welfare plans provide non-pension benefits, most importantly health, accident, disability, and life insurance benefits. 29 U.S.C. § 1002(1) (2000).

⁷⁹ 29 U.S.C. § 1002 (2006) (defining plan assets).

Department of Labor had to supply a more appropriate meaning through a modestly complex regulation.⁸⁰ Further, to secure the safety of plan financial resources, ERISA subjects parties who control and manage the financial resources of employee benefit plans to fiduciary duties.⁸¹ These duties are stated in three or four relatively short statutory provisions; in comparison, the standard treatises devote dozens of pages to explaining fiduciary duties.⁸² There are, in addition, thousands of cases decided at common law and under state trust codes explicating and refining fiduciary duties. And, of course, it is not always possible to harmonize these cases. Regarding fiduciary duties and the nuances those duties entail, ERISA is textually highly abbreviated and incomplete. Accordingly, to create a body of fiduciary law that is not set out in the text of this "comprehensive and reticulated statute," the federal

⁸⁰ 29 C.F.R. § 2510.3-101(2004). One of ERISA's purposes is to protect the assets of employee benefit plans, especially from mismanagement and self-dealing by fiduciaries. However when the phrase "plan assets" is understood in terms of the common law meaning of "asset," the Act contains a substantial gap in protection. Plan assets can be converted into non-plan assets, under the common law definition, and yet be the economic equivalent of plan assets. For example, the assets of a corporation are not the assets of the shareholders. Manipulative parties can readily convert plan assets into interests in entities, the assets which are not assets of the plans in traditional parlance. Yet, the assets of the entities can economically and practically be the equivalent of assets of the plans that invest in them. It was, therefore, necessary to refine, by expansion, ERISA's purpose of protecting plan assets by redefining the term "plan assets" to include the underlying assets in certain entities in which employee plans invest. Another way of stating the result is to say that one of ERISA's purposes was not simply to protect financial resources contributed to plans while such resources remained in the form of common law assets, but rather to provide protection even when such resources are converted into other forms as long as it makes sense to view them economically and practically as substantially equivalent to common law assets.

⁸¹ For purposes of ERISA a party is a fiduciary with respect to an employee benefit plan, *inter alia*, to the extent the party "exercises any discretionary authority or discretionary control respecting management of such plan or exercises any authority or control respecting management or disposition of its assets" 29 U.S.C. § 1002(21)(A) (2000).

⁸² Compare, for example, the scope of the treatment of fiduciary duties in 29 U.S.C. §§ 1104 -1106, 1109 (West 1999), with the treatments of fiduciary duties in GEORGE GLEASON BOGERT, GEORGE TAYLOR BOGERT & AMY MORRIS HESS, *THE LAW OF TRUSTS AND TRUSTEES* (rev. 2d ed. 1984); AUSTIN WAKEMAN SCOTT & WILLIAM FRANKLIN FRACHTER, *THE LAW OF TRUSTS* (4th ed. 1987).

courts have been required to develop what is in fact a potentially substantial "federal common law of trusts."⁸³

ERISA also requires that all plan assets be held either in trust or in insured arrangements.⁸⁴ Many employers elect to have the plans they sponsor hold all or part of their assets in trust. However, through a provision that was added to the act at the last minute and without any time for deep—or even shallow, for that matter—reflection on its ultimate consequences, ERISA preempts all state laws that relate to employee benefit plans.⁸⁵ The purpose of the preemption provision is not only to protect the federal statutory scheme from state infringement but also to ease plan administration by insuring uniformity of the law applicable to employee benefit plans.⁸⁶ State trust laws are clearly among the state laws that relate to employee benefit plans that hold plan assets in trust. Since state trust laws may prohibit conduct that ERISA permits and may permit conduct that ERISA prohibits, state trust laws are clearly one of the types of state law that ERISA was intended to preempt. However, in spite of the fact that ERISA requires plan assets to be held in trust and simultaneously preempts all state trust laws, ERISA does not contain a federal trust code. Rather, it contains little more than one line requiring that plan assets be held in trust. That is a substantial gap, or omission, in the substantive scheme of the law. That also has been one of the impelling motives behind the development of a federal common law of trusts.

Further, although one of the paramount purposes underlying ERISA's fiduciary duties was the protection of plan financial resources, ERISA does not contain any general substantive rules governing the conduct of non-fiduciaries who may act in concert with plan fiduciaries to cause losses of plan assets. Neither does ERISA's text contain any general rules

⁸³ See, e.g., *Firestone Tire & Rubber Co. v. Bruch*, 489 U.S. 101 (1989).

⁸⁴ 29 U.S.C. § 1103 (2000).

⁸⁵ 29 U.S.C. § 1144 (2000), amended by Pub. L. No.109-208, 120 Stat. 780 (2006).

⁸⁶ See, e.g., *Fort Halifax Packing Co., Inc. v. Coyne*, 482 U.S. 1 (1987); *Moore v. Philip Morris Cos., Inc.*, 8 F.3d 335 (6th Cir. 1993).

governing the conduct of those who provide services to plans but are not fiduciaries.⁸⁷ Since ERISA preempts all state laws that relate to employee benefit plans, it would appear, in light of ERISA's text alone, that those who aid and abet plan fiduciaries in breaches of duty and non-fiduciary professionals, such as lawyers, accountants, and actuaries, who engage in malpractice are at liberty to do so.⁸⁸ Courts, not surprisingly, have on occasion intervened and effectively created substantive rules and remedies in spite of the absence of statutory text.⁸⁹

Of course, since ERISA contains no general substantive rules addressing the conduct of non-fiduciaries, it also contains no remedial provisions regarding the conduct of such persons. So in the case of non-fiduciaries, ERISA lacks both substantive and remedial provisions.⁹⁰

The above are only a miniscule fraction of the issues for which the text of ERISA fails to provide a rule.⁹¹ The degree

⁸⁷ Certain parties, called "disqualified persons" in I.R.C. § 4975(e)(2) (2002) and "parties in interest" under § 3(14) of ERISA, 29 U.S.C. § 1002(14) (West 1999), who may not be fiduciaries, are subject to excise taxes for engaging in various categories of what are called "prohibited transactions." I.R.C. § 4975(a)-(c) (West 2002). There is, however, no provision in ERISA that subjects non-fiduciaries to liability to employee benefit plans.

⁸⁸ The failure to provide substantive rules prohibiting the actions of aiders and abettors of fiduciary breaches and providing for malpractice liability for non-fiduciary professional service providers to employee benefit plans might appear to be an odd omission until one reflects on the fact that one of the driving forces impelling the enactment of ERISA was a series of serious breaches of fiduciary duty by both union and corporate sponsors of plans. The problems posed by aiders and abettors and professionals engaging in malpractice were simply nowhere on the radar screen either when the substantive provisions of ERISA were being crafted or when ERISA's preemption clause was added at the last moment before passage.

⁸⁹ See *infra* note 155.

⁹⁰ In contrast, the Securities Exchange Act of 1934, 15 U.S.C. §§ 78a-78mm (2000), contained clear prohibitions on fraud in connection with the sales and purchases of securities, but it contained no remedies for fraud. The courts, not surprisingly, found an implied remedy. *Kardon v. Nat'l Gypsum Co.*, 73 F. Supp. 798, 800 (E.D. Pa. 1947).

⁹¹ Congress recognized with respect to ERISA that it was producing an incomplete rule scheme. The statute, accordingly called for many regulations. In such cases there was a conscious recognition by Congress of specific gaps in the statu-

to which even a supposedly "comprehensive and reticulated" statute is incomplete is evidenced in the difference between ERISA's approximately 220 pages of none too compact initial text⁹² and the rules in that text and the thousands of pages of text and even greater number of rules contained in the aggregate of the Treasury Department Regulations (877 pages of small, double column print that dwarf the original statutory provisions) relating to the act⁹³ (not all of which have been promulgated even after thirty-one years) and useful substan-

tory rule system. When it is highly obvious that this is so, Congress often expressly delegates authority to a regulatory agency to promulgate regulations designed to render the law more complete. Of course, Congress does not always recognize the gaps, or does not always expressly recognize them, and the need to deal with these matters lands in the lap of the courts. The task facing the courts and the regulatory agencies is essentially the same. Since even the regulatory agencies cannot thoroughly complete the law, the courts always end up with the residual cases for which no one else, Congress or the regulatory agencies, has provided a result.

Some of the regulations under ERISA were required to address questions of commencement of participation in employer sponsored pension and welfare plans, the counting of service under plans, accrual of plan benefits, actuarial requirements for pension plan funding, vesting in benefits under pension plans, calculation of employer liabilities for withdrawal from multi-employer pension plans, the calculation of payments to the benefit insurance program to be established by the Pension Benefit Guaranty Corporation, and, under the Internal Revenue Code's qualification provisions, guidelines designed to prevent discrimination in either contributions or benefits under pension plans in favor of highly compensated employees. The regulations regarding pension plan benefit accrual and funding, multi-employer plan withdrawal liabilities, and the discrimination of benefits in favor of highly compensated employees had to include complex actuarial guidelines, matters far beyond the expertise of the average member of Congress. And the authorized regulations were not the end of the story. The regulatory agencies have had to establish many additional rules through interpretive regulations that the agencies hoped the courts would uphold as valid interpretations of the statute.

⁹² The original text of ERISA, Pub. L. No. 93-406; 86 Stat. 829, required 252 pages in 1974 U.S.C.C.A.N., (86 Stat.) 935. A number of pages of text are duplicated in that a number of labor law and tax rules in Titles I and II are mirror image provisions. Eleven pages of the statute concern agency administrative matters, and some additional material does not state rules but specifies where in existing laws amendments should be inserted. Altogether, the actual, non-duplicated rules probably constitute around 220 pages.

⁹³ The overly curious can find the majority of these regulations in 26 C.F.R. §§ 1.401-1.420-1 (2004).

tive information in the preambles to those regulations; Treasury Department Revenue Rulings, private letter rulings, Revenue Procedures, General Counsel Memoranda, Technical Advice Memoranda, and the like; Department of Labor and Pension Benefit Guarantee Corporation Regulations (656 pages of small, double column print that dwarf the original statutory provisions) relating to the act⁹⁴ (not all of which have been promulgated even after thirty-one years) and useful substantive information in the preambles to those regulations, Department of Labor and Pension Benefit Guaranty Corporation Advisory Opinions and Opinion Letters; over twenty volumes of trial and appellate court opinions;⁹⁵ and scholarly commentary addressing matters not resolved by all of the foregoing, even when all duplicative and unnecessary text in those various sources is disregarded. And the resulting aggregate thousands of pages of text and even more numerous rules contained therein still do not come close to addressing all questions that arise in the course of practice in the employee benefits arena. That is, in part, because regulations, rulings, opinions letters, and cases always raise questions as well as provide answers. There are thousands of issues relating to this "comprehensive and reticulated" statute arising out an unlimited number of unanticipated, sometimes unique, circumstances that have not yet been authoritatively addressed.

It is obvious that even approaching a truly complete Employee Retirement Income Security Act, given the driving purposes underlying the act, would—in fact, does—require many thousands of pages of highly compact text. If society had been required to wait for the statute's protections until the statute had been rendered textually complete, there would never have been an ERISA.⁹⁶ Indeed, the best that Congress

⁹⁴ The overly curious can find the majority of these regulations in 29 C.F.R. §§ 2509.75-2-4907.170 (2003).

⁹⁵ See the Employee Benefit Cases volumes published in connection with the BNA Pension & Benefits Reporter. Even that collection of ERISA cases is not exhaustive, as the author has learned by running Lexis and WestLaw searches.

⁹⁶ That would fulfill with a vengeance Thomas Jefferson's belief that the government that governs best governs least. Were Congress to wait until it could

could do, given the constraints of time under which it operated, was to provide many essentials and rely on the regulatory agencies, practitioners, and courts to navigate to many desired results that could not be addressed *ex ante*. That is the case, in greater or lesser degree, with every statute. As things stand, ERISA, rather ironically, stands for the proposition that the more "comprehensive and reticulated" a statute is, the more the work that will fall on administrative agencies and courts in rendering it substantially complete.

C. "Typical" Statutes

In between the statutes above are those statutes that no one has ever viewed as either "common law" or "comprehensive and reticulated" statutes. As rule systems, these statutes are all in significant ways incomplete. Consider, for example, Title VII of the Civil Rights Act of 1964 ("Title VII").⁹⁷

The purposes of Title VII are to discourage discrimination against individuals in hiring or with respect to any term or condition of employment on the basis of race, color, religion, sex, or national origin and to compensate individuals who have suffered discrimination.⁹⁸ While it may be conceptually mandatory that employers be parties liable for discrimination, once that decision has been made it is necessary to decide, *inter alia*, who is an employer, who is an employee or potential employee,⁹⁹ and whether one employer is to be held liable

provide complete statutory schemes covering all possible contingencies, it would, in fact, govern not at all.

⁹⁷ 42 U.S.C. §§ 2000e-2000e-17 (2000).

⁹⁸ 42 U.S.C. § 2000e-2(a)(1) (2000).

⁹⁹ The determination of who is to be treated as an employee is not a simple matter. Because of the different purposes underlying various labor and employment related legislation courts have had to tailor the definition of "employee" to render it appropriate to various acts in which it appears. These definitions are in large part pragmatically justified judicial contributions to the American English lexicon. *See, e.g.*, *NLRB v. Hearst Publ'ns*, 322 U.S. 111 (1944) and *Walling v. Portland Terminal Co.*, 330 U.S. 148 (1947) (the court held that the word "employee" for purposes of National Labor Relations Act, 29 U.S.C.A. § 152 (2000), and the Fair Labor Standards Act, § 203(e), means, generally, common law employees and certain individuals who satisfy criteria (in part different under each

for causing another to discriminate.¹⁰⁰ The Act contains no rules addressing these issues. A complete rule system would also either exhaustively define or catalogue all the acts that constitute discrimination.¹⁰¹ Congress made no attempt to do that. In addition, although it was decided to limit liability to employers as entities, the Act did not specify all the circumstances in which an employer is deemed to act. Since a corporation, for example, always acts through agents, a complete rule system would necessarily specify when a corporation is to be liable for which acts of which of its agents.¹⁰² There was no attempt to specify whether all employees are agents of their employers for purposes of Title VII, or whether only limited classes of employees are deemed to be agents.¹⁰³ Further, the Act did not specify whether failures by employees

act) that would generally render individuals independent contractors.) Justice Jackson pointed out that the common law test for employees was devised for the primary purpose of "deciding whether one who hires another is responsible in tort for his wrongdoing." *Hearst Pub'ns*, 322 U.S. at 120. That purpose was significantly different from the purposes underlying the acts at issue, and the Court made modifications in the concept of "employee" to bring it into line with the purposes of the acts. *A priori* there is nothing to lead one to believe that a classification appropriate to the allocation of enterprise liability would be perfectly suited to other legal objectives.

¹⁰⁰ See, e.g., the case annotations to 42 U.S.C.A. § 2000e (West 1994) nos. 107-166 and the cases addressing, e.g., whether and when two employers are to be deemed a single employer and who is deemed to have actual control over employment.

¹⁰¹ The courts were required to determine, e.g., whether facially neutral employment criteria having adverse impact on a protected class, managerial inactivity in the face of harassment of a protected class, and same sex harassment constituted "discrimination." See *Oncale v. Sundowner Offshore Servs., Inc.*, 523 U.S. 75, 79 (1998) (same sex harassment is actionable); *Harris v. Forklift Sys., Inc.*, 510 U.S. 17, 21 (1993) (harassment is to be judged from the perspective of a "reasonable person"); *Meritor Sav. Bank, FSB v. Vinson*, 477 U.S. 57, 72-73 (1986) (harassment of a protected class that an employer allows can constitute discrimination); *Griggs v. Duke Power Co.*, 401 U.S. 424, 430 (1971) (holding that facially neutral employment criteria that have an adverse impact on a protected class are actionable unless job related and consistent with business necessity, terms hardly paragons of precision).

¹⁰² See, e.g., *Burlington Indus., Inc. v. Ellerth*, 524 U.S. 742 (1998); *Faragher v. City of Boca Raton*, 524 U.S. 775 (1998).

¹⁰³ *Id.*

and other agents to act are to be treated as acting.¹⁰⁴ Nor did the Act specify whether all employees and agents are deemed the same for all purposes of the foregoing determinations.¹⁰⁵ A complete rule system addressing those matters would have required supplying agency and vicarious liability rules consonant with the answers. One might expect that the answers and rules would vary depending on the type of discriminatory conduct at issue and who is engaging in it.¹⁰⁶ No such rules were supplied.

One of the more perplexing problems regarding Title VII simultaneously involved the substance of the Act, the law of evidence, and the rules of procedure. It turned out that Title VII required its own special rules of evidence and procedure, which, in turn, highlighted a need for a substantive addition. The problem was how to prove discrimination when an employer employed facially neutral employment criteria that excluded members of a protected class from employment. That is, suppose an employer requires all successful job applicants to hold a high school diploma. Suppose, further, that many of the jobs the employer holds out, such as sweeping and cleaning, do not require the level of education associated with a high school diploma. And, then, suppose finally it turns out that a smaller proportion of a protected minority than of the majority in the relevant job market holds high school diplomas.¹⁰⁷ Proving discrimination in such situations required resort to statistical methods (in no way even hinted at in Title VII), whereby a sufficient disparity between the qualified members of a protected class employed by a given employer and the available pool of qualified members of the protected class in the relevant labor market was deemed sufficient evidence of discrimination to establish a *prima facie* case.¹⁰⁸

¹⁰⁴ *Id.*

¹⁰⁵ *Id.*

¹⁰⁶ As Lance Liebman, Director of the American Law Institute, observes, special rules of agency have evolved in employment law. Lance Liebman, *Foreword* to RESTATEMENT (THIRD) OF AGENCY (Tentative Draft No. 1, 2000).

¹⁰⁷ *Griggs v. Duke Power Co.*, 401 U.S. 424, 430 (1971).

¹⁰⁸ See *id.* at 430-32; *Dothard v. Rawlinson*, 433 U.S. 321, 330 (1977);

The courts decided that when such a case was established, the employer should be permitted to rebut it by showing that the employment criteria were "job related and consistent with business necessity."¹⁰⁹ The substantive fall out was that what might be truly unintended discrimination, negligent discrimination, if you will, was declared illegal along with intentional discrimination.¹¹⁰

Many of the matters above that were not addressed by the textual rule system of Title VII have, of necessity, been filled-in by the courts. There are, of course, hundreds of additional matters that have had to be decided judicially to further the process of completing Title VII's rule system. A run through the annotations to Title VII in the United States Code Annotated will give a nice sample, although those cases are not exhaustive, and there are state laws with language identical or similar to that in Title VII that address matters not yet addressed under Title VII. Admittedly, not every case annotated truly provides a missing rule. Some are duplicitious. Still, the number is not trivial.

In addition to the cases, of course, there are the Title VII regulations, many being interpretive, addressing a variety of matters that the text of the statute omitted.¹¹¹ And after nearly forty years the need for additional rules continues to surface. We apparently have not yet encountered in reported cases every single behavior and combination of facts for which a rule must be established if the purposes of Title VII are to be effected. There is no reason to believe we ever will.¹¹²

Albermarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975). The "disparate impact" theory of discrimination is codified in 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2000).

¹⁰⁹ 42 U.S.C. § 12113 (2000).

¹¹⁰ *Griggs*, 401 U.S. at 424.

¹¹¹ See 29 C.F.R. 1601.1-1604.11, 1605.1-3, 1606.1-8, 1607.1-8, 1608.1-12 (2005).

¹¹² Regarding inherent incompleteness, statutory rule systems and common law rule systems are quite similar. The primary differences are, inter alia: (1) the relevant intentions and purposes in statutory rule systems are those of the legislature and in the common law rule systems those of the courts, which hopefully reflect community values, (2) it more often takes courts an extended period of time to determine and refine the purposes they want to pursue, and (3) common

If we move from illustrative anecdotes to system, we find there are at least sixteen types of ways in which statutes can emerge from the enactment process incomplete. The first four relate to the purposes of statutes.

First, although, as noted above, cognitive science, pragmatics, epistemology, and the philosophy of the mind make it clear that the purpose or purposes of a statute are among the most essential elements in understanding the human behavior evidenced in a statute and the behavior it is supposed to induce, statutes occasionally emerge with little or no articulation of purposes.¹¹³ Further, no individual or group of individuals, whether legislators or legislative staffs, has the intelligence, prescience, creative imagination, or time, *ex ante* to uncover or anticipate all the possible clusters of actions, behaviors, events, circumstances, relations, consequences, and the like "facts" that reality can present over time. As a result, the purposes of a statute, *i.e.*, the results it has been designed to achieve, not only occasionally emerge with little or no articulation, or only vague articulation, they occasionally emerge in part inchoate,¹¹⁴ almost always, if not always, without full specification of scope, and almost always without fully refined lexical statements. The issues presented

law rule systems begin life more incomplete than do statutory rule systems. Courts have traditionally been loathe to formulate their own complex, detailed rule systems, with manifold exceptions and special modifications to the meanings of words, absent issues in actual controversy. Courts are also reluctant to address matters in anticipation of future litigation. Common law courts use the principle of holding to limit the scope of their opinions so that they retain freedom to deal with overlooked and unanticipated matters in light of the policies giving rise to their rule systems. Legislators, in contrast, attempt to anticipate and address *ex ante* a myriad of matters in a statute that they believe will or are likely to arise. But the differences between statutory and common law rule systems are matters of degree. And, as long as courts pursue consistent purposes in connection with the development of common law rule systems, anticipation of future developments in the common law and the interpretation of statutes are virtually identical processes.

¹¹³ Both the Sherman Act, 15 U.S.C. §§ 1-7 (2000), and the Clayton Act, 15 U.S.C. §§ 12-17, omit recitations of purposes. This is explained *supra* text at pp. 79-84.

¹¹⁴ See, *e.g.*, the Sherman Act, 15 U.S.C. §§ 1-7 (2000).

by the last three items are not self-identifying. Rather, the issues become apparent only when overlooked or unanticipated facts surface and reveal that the reasonably ascertainable purposes of a law, as initially formulated, are in need of significant completion, clarification of scope, or lexical refinement.

As a practical matter, a statute's purposes must be initially formulated on the basis of a limited set of potential behaviors, occurrences, and the like. These behaviors, occurrences, and the like are what bring a problem requiring legislative action to the attention of interested parties. These behaviors, occurrences, and the like can be called "paradigmatic facts." The purposes of the statute will, inevitably, be conceptualized and stated in terms of the paradigmatic facts. In the case of some statutes the paradigmatic facts are an exceptionally small subset of the entire set of facts that need to be addressed through the statute. That was the case with the Sherman Act. With micro-economics in its infancy, in fact, a mere new born on the date of enactment, and the full implications of the second industrial and accompanying financial and institutional revolutions only dimly perceived, no one could fully understand the problems that the Act would need to address. Accordingly, the purposes of the Sherman Act emerged partially inchoate.

The Sherman Act represents a somewhat rare extreme. Much more commonly, a statute covers a significantly complete set of paradigmatic facts.¹¹⁵ In virtually all such instances, it is, nonetheless, necessary over time to refine the scopes of a statute's purposes.

When a statutory purpose is initially formulated, its scope is likely to be textually either too narrow or too expansive or both. Suppose, for example, a statute is designed, *inter alia*, to prevent the behaviors evidenced in a given set of paradigmatic facts because those behaviors cause injury that public policy demands be curtailed. The behaviors present in the

¹¹⁵ That is because a great deal more is known about the underlying subject matter to which the law is to apply than was the case with the Sherman Act.

paradigmatic facts may not, however, be the entire set of behaviors that can give rise to the injury in question. There may be overlooked and unanticipated behaviors that can be called "non-conceptual equivalents" of the behaviors present in the paradigmatic facts,¹¹⁶ behaviors that, although not conceptually identical to and possibly quite conceptually distinct from the paradigmatic behaviors,¹¹⁷ can, nonetheless, give rise to the injury of concern. Such non-conceptual equivalents may not be surfaced at any point during the formulation of the statute's rule system. In fact, non-conceptual equivalents may not even exist when a statute is enacted but, rather, may be concocted long after the statute is enacted as parties, with the assistance of counsel, creatively seek ways to evade its strictures. Lawyers are often highly rewarded for concocting schemes for accomplishing results that statutes are designed to prevent. They do so by finding alternative ways of accomplishing those results that are not conceptually captured by the words and phrases used in statutory rules.¹¹⁸ To the ex-

¹¹⁶ We often encounter non-conceptual equivalents in the law. For example, a corporate merger can be the equivalent of a simple amendment to the articles of incorporation; a common law trust can be designed as the equivalent of a holding company; a trust certificate can be the equivalent of a stock certificate; a corporate leveraged buyout can be the equivalent of an illegal dividend distribution creating an illegal, fraudulent preference or asset distribution in favor of stockholders over unsecured creditors. *See In re Bay Plastics, Inc.*, 187 B.R. 315, 333-35 (Bankr. C.D. Cal. 1995). A Chapter 13 bankruptcy proceeding can produce results equivalent to a Chapter 7 proceeding that cannot be instituted because the debtor has secured a Chapter 7 discharge within the prior six years. *See, e.g., In re Strauss*, 184 B.R. 349, 351 (Bankr. D. Neb. 1995). A series of tax free reorganizations can be equivalent to a dividend distribution. *See, e.g.,* the discussion of the step-transaction doctrine, BORIS I. BITKER & JAMES EUSTICE, *FEDERAL INCOME TAXATION OF CORPORATION AND SHAREHOLDERS*, ¶¶ 1.05[2][d], 12.61[3] (6th ed. 1994); Marvin A. Chirelstein & Benjamin B. Lopata, *Recent Developments in the Step-Transaction Doctrine*, 60 TAXES 970 (1982); Joseph Isenbergh, *Musings on Form and Substance in Taxation*, 49 U. CHI. L. REV. 859 (Summer 1982). Also, it is possible through contractual arrangements to create an economic entity that is equivalent to what can be achieved through merger. *See, e.g., Rotherby Storage & Van Co. v. Atlas Van Lines, Inc.*, 792 F.2d 210 (D.C. Cir. 1986), *cert. denied*, 107 S. Ct. 880 (1987).

¹¹⁷ To take one example from the immediately preceding footnote, a corporate merger and an amendment to the articles of incorporation are conceptually distinct, although the former can be a vehicle for the latter.

¹¹⁸ Occasionally, a particular way of achieving a given result, not the result, is

tent non-conceptual equivalents are not recognized or anticipated and included within a statute's coverage, the scope of the statutory purposes relating to those non-conceptual equivalents will be too narrow and the statute will not textually reach as far as it should.

For example, clever tax lawyers discovered that by properly sequencing a series of "tax free" corporate reorganizations and following up with a corporate liquidation, one could effectively distribute corporate earnings and profits, taxable as ordinary income, in what was in form a sale or exchange of a capital asset, thereby miraculously converting ordinary income into capital gains.¹¹⁹ Although a series of corporate reorganizations followed by a corporate liquidation is conceptually distinct from a dividend declaration and payment, the economic effect can be identical. Avoidance of ordinary income tax through legal manipulations that have no economic substance was never what the tax free reorganization provisions were intended to allow. That was implicit, albeit not express, in the in totality of the scheme for corporate taxation.

But there is an additional problem. Assuming again that the statute at issue exists to prevent or discourage an injurious form of behavior, there may be injuries that are so similar to—even if not in every respect identical to—those evidenced in connection with the paradigmatic facts and non-conceptual equivalents that no rational actors would act to prevent the latter without also acting to prevent the former.¹²⁰ The be-

what creates a concern. In such instances, the results may be quite socially useful. When that is true, lawyers who assist in the creation of alternative ways of achieving the results that do not give rise to the original concerns are true problem solvers rather than simple manipulators.

¹¹⁹ BITKER & EUSTICE, *supra* note 116; Chirelstein & Lopata, *supra* note 116; Isenbergh, *supra* note 116.

¹²⁰ Thus, although the Fair Labor Standards Act's provisions for minimum wages was based on a paradigmatic concern with the wages of common law employees, individuals over whom an employer exercises the power to hire and fire, usually at will, and with respect to whom an employer dictates not only the job to be performed, but how, when, and where to perform it (and usually supplies the place and equipment for work), it soon became clear that independent contractors required the same protections even though the relations between independent contractors and the employers with whom they contract are significantly

haviors that give rise to these injuries may be called "substantial equivalents."¹²¹ Some or all of the substantial equivalents may also be overlooked or not anticipated during the formulation of the statute. Further, as is the case with non-conceptual equivalents, parties may devise substantial equivalents well after the statute is enacted.

Since, absent highly unusual circumstances, the intention of the parties formulating a statute, were they to uncover or anticipate them, would virtually always be to address the aggregate of paradigmatic facts and non-conceptual and substantial equivalents, an intention to cover non-conceptual and substantial equivalents along with the paradigmatic facts must rationally be presumed. No rational, honest actor would intentionally cover the paradigmatic facts but omit coverage of non-conceptual and substantial equivalents. When non-conceptual and substantial equivalents are omitted due to oversight or the failure to or inability to anticipate them, the statute's purposes, as ascertainable from text alone, will be too narrow, and the statute will not reach as far as it should and would but for oversight and the failure or inability to anticipate. The scope of its purposes ascertainable from text alone will be too narrow.

Of course, the scope of statutory purposes can also be textually too expansive. Even if all behaviors giving rise to paradigmatic facts and non-conceptual and substantial equivalents were identified and covered by the statute's strictures, those facts could occur with other facts in light of which it

different from the relationship between employer and employee. Also, to preclude employers from converting employees to independent contractors on every occasion on which it might be feasible to do so, it became necessary to treat all persons who render services to an employer or are suffered to render services as in the same class for purposes of the minimum wage protections. See 29 U.S.C. § 203(d)-(g) (2000).

¹²¹ Antitrust law is replete with substantial equivalents. Price fixing and territorial division, for example, are effective ways to achieve the benefits of monopoly without actually achieving true monopoly status. Further, there are several substantial equivalents to direct price fixing, one being the sharing of otherwise proprietary cost, production, and pricing information so that producers can engage in "conscious parallelism" and tacit agreements in their pricing.

would be irrational, pointless, or unfair to impose the statute's strictures. These facts may be called "exceptional fact aggregates." For example, imposing the strictures of a statute in the face of the additional facts could on occasion produce results contrary to the very purposes of the statute.¹²² It is also possible that imposing statutory strictures, although not yielding a result contrary to the purposes of the statute, would, nonetheless, not only fail to advance the purposes of the statute but would serve no sound purpose at all.¹²³ In these in-

¹²² Thus, the owners or managers of an enterprise competing in a market, without significant barriers to entry and purely on the merits through continuous innovation and superior service, and in no way abusing its financial, marketing, or other power, and refraining from anti-competitive mergers and acquisitions, may intend to drive all the competitors of the business out of the market and become the sole supplier in that market, and in fact do so. Such conduct would not generally be viewed as subject to the anti-monopolization strictures of the anti-trust laws because to so class it would dampen what the anti-trust laws are designed to do—foster and protect competition on the merits. If successful, such a competitor, although in one sense a monopolist, i.e., sole economic actor in a given market, would also be the perfect competitor, operating so efficiently, keeping costs and prices so low, innovating so effectively, and commanding such customer loyalty through superior products and services that, in spite of the absence of significant barriers to entry, no one would want to attempt to enter the market. The enterprise also would not earn monopoly profits.

¹²³ This appears to be one of the rationales that gave rise to the decision in *Rector of Holy Trinity Church v. United States*. 143 U.S. 457 (1892). Congress had made it illegal to "in any way assist . . . the importation . . . of any alien . . . into the United States, . . . under contract or agreement . . . made previous to the importation . . . of such alien, . . . to perform labor or service of any kind in the United States . . ." Act of February 26, 1885, ch. 164, § 1, 23 Stat. 332. The Episcopal Church of the Holy Trinity entered into an agreement with an Anglican minister residing in England pursuant to which the minister would immigrate to the United States and become the minister of the church. The statute in question literally covered the transaction. However, the statute was clearly designed to protect the American working class from the wage competition and wage depression resulting from the massive importation of cheap foreign labor. "All the supporters of the bill who spoke during the floor debates saw its purpose to be preventing the importation of 'laborers' who would undermine the wage position of American wage earners, the precise purpose articulated in the committee reports and in the Supreme Court opinion." William N. Eskridge, Jr., *Textualism, The Unknown Ideal?*, 96 MICH. L. REV. 1509, 1538-39 (1998) (footnote omitted). There was, however, no such problem presented in the case before the Court. Anglican ministers were not in excess supply. As the Court noted: "It was never suggested that we had in this country a surplus of brain toilers, and, least of all, that the market for the services of Christian ministers was depressed by

stances, if the law were, for example, punitive, application of the law would impose a penalty that senselessly harms.¹²⁴

foreign competition." *Holy Trinity Church*, 143 U.S. at 464. The Court, accordingly, held that the importation of the minister was not covered by the statute's prohibition. This logic is recognized in the current H1B visa program pursuant to which employers can petition to employ foreign nationals who have unique qualifications that cannot be satisfied out of the available pool of talent in the United States. For opposing viewpoints regarding the decision in *Holy Trinity Church*, see for example, SCALIA, *supra* note 4, at 18-23; William N. Eskridge, Jr., *Textualism, The Unknown Ideal?*, 96 MICH. L. REV. 1509 (1998). In the author's view, the statute in question represents a case of what may be called "protectively overly broad drafting," drafting that may not be viewed literally without close regard to underlying purposes. This technique is used when the drafting parties cannot identify the precise category or categories they want to cover and are concerned that, due to the tendency toward literalness displayed by some judges, some "fact types" they want to cover but cannot ex ante identify may escape coverage. They, therefore, select words or phrases that sweep so broadly that all matters that should be covered, given the purposes of the law, will virtually certainly be encompassed. It is the responsibility of intelligent interpreters to determine when the statute's purposes do not require coverage. For possibly the most famous, or infamous, example of protectively overly broad drafting, see *infra* pp. 116-18 and note 154.

¹²⁴ Thus, the Occupational Safety and Health Act of 1970, 29 U.S.C. §§ 651-700 (2000), the purpose of which is to provide workers with safe and healthful workplaces, § 651(b), requires employers, inter alia, to comply with all occupational safety and health standards promulgated by the Secretary of Labor. § 654(a)(2). The Act provides that those who fail to comply are subject to monetary penalties and must abate non-conforming conditions. § 666. The Act, itself, however, also provides that "[t]he Secretary may prescribe procedures for the issuance of a notice in lieu of a citation with respect to de minimis violations [of promulgated safety and health standards] which have no direct or immediate relationship to safety or health." § 658(a). The Occupational Safety and Health Administration has provided standards for notices of de minimis violations in its Field Inspection Reference Manual CPL 2.103, Section 7, Chapter III, at Subsection C.2.g. (July 13, 1999). Such notices are to be issued, inter alia, when:

- (1) An employer complies with the clear intent of the standard but deviates from its particular requirements in a manner that has no direct or immediate relationship to employee safety or health.
- (2) An employer complies with a proposed standard or amendment . . . rather than with the standard in effect at the time of the inspection and the employer's action clearly provides equal or greater protection
- (3) An employer's workplace is at the "state of the art" which is technically beyond the requirements of the applicable standard and provides equivalent or more effective employee safety or health protection.

Or, the benefits of imposing the strictures could be de minimis yet grossly unfair and unjust. Further, imposing the strictures could further the statute's purposes in a significant way but, nonetheless, run counter to an overriding purpose or set of purposes, including those evidenced in other statutes or the Constitution.¹²⁵

The practical matter is that if exceptional fact aggregates are overlooked or not anticipated while a statute is being formulated, there will be no textual exceptions provided for them. When exceptional fact aggregates are not textually excluded, the statute will be purposively incomplete in that it will reach further than actually intended. And, of course, there can be facts that should be treated as exceptions to these exceptions, facts that may be called "extra-exceptional fact aggregates." If these are not addressed, but exceptional fact aggregates are, the statute will fail to extend as far as it should.¹²⁶

While addressing non-conceptual and substantial equivalents, exceptional fact aggregates, and extra-exceptional fact aggregates often requires additions to text (on many occasions these matters can be addressed by definitional modifications), other gaps in legal rule systems can also require effective additions to text. As noted above with respect to both ERISA and Title VII, statutes sometimes emerge lacking entire subsidiary rule systems. In addition, given the presumptive semantic content of statutory rules, observant, skillful, and creative lawyers often find applications for rules that were

De minimis violations do not give rise to either monetary penalties or an obligation to abate. The logic is that, in such circumstances, penalties and abatement, i.e., compliance with specific safety and health standards, would not advance the purposes of the law and would, therefore, be pointless.

¹²⁵ To lean on common law for an analogy, the privilege of self-defense in battery cases overrides the general prohibition on intentionally striking others in harmful ways. This principle is also obviously operating in *Riggs v. Palmer*, 22 N.E. 188 (1889), in that the statutory prohibition of murder precludes allowing murderers to benefit from their illegal conduct, since doing so would tend to encourage murder by rewarding it.

¹²⁶ See *infra* note 161.

almost certainly neither anticipated nor contemplated by the parties formulating the rules.¹²⁷ Or, they may discover ways to achieve results through the sequential application of rules that the party formulating the rules never contemplated.¹²⁸ Since these uses were not anticipated, there are no rules prohibiting them. Nonetheless, those uses may run counter to the statute's purposes or overriding purposes—purposes that should prevail because the applications in question were not contemplated, foreseen, or intended, but result from the mere happenstance or accident that the rules can literally be so used.

The fact is that, except in the rarest circumstances, if any, it is not humanly possible in devising a legal rule system to be aware of or anticipate every possible simultaneous occurrence of facts and provide a rule for every eventuality. There are potentially an infinite number of real world situations, many only occurring rarely, that are not dealt with by any statutory rule system as it first emerges from the legislative processes. Were all potential situations addressed, many laws would be infinitely long sets of rules, with an unlimited number of special situation additions and exceptions. Accordingly, the initial formulation of a statute, even if the purposes are in theory fully clarified in the minds of the parties formulating them, can never be more than a partial statement of the rule system for achieving the purposes of the statute and all simultaneously existing purposes that impinge on the statute. The partial

¹²⁷ Examples include directors' control over corporate distributions to squeeze out minority shareholders at a loss; the use of mergers to effect amendments to the articles of incorporation that are otherwise blocked by the existing articles of incorporation, or to expel a minority interest; the use of a dividend to avoid a charge to the income statement, *Kamin v. American Express*, 383 N.Y.S.2d 807 (1976); the use in the late Nineteenth Century of trusts as vehicles for business combinations not possible under then existing corporate law, absent special legislative grace, which was not always forthcoming or was too costly in bribes; trust certificates as substitutes for stock certificates; LLC's, the formation of which is not limited to "for profit" activities, as asset protection devices; and the formation of LBOs as vehicles for preferential distributions to owners in relation to unsecured creditors. Tax law and corporate law practices, in particular, involve many creative—never before tried/never before ruled on—applications of rules.

¹²⁸ See *supra* note 116.

statement may be quite good, addressing most of the situations that actually occur in the world, but it will require judicial completion both consistent with the purposes, and directed at fulfilling and furthering those purposes, as rarely occurring situations arise.¹²⁹

To recapitulate and complete the observations made here in Part III, statutory purposes emerge incomplete, or unrefined, in several, if not all, of the following four ways:

1. Although the purposes of the statute may be well understood by the parties bringing it into existence, those purposes may not be stated or clearly or fully articulated in text.
2. In some instances, the purposes of the statute are, in part, highly inchoate because those who bring the statute into existence have not fully formulated and do not themselves fully understand its purposes.
3. In virtually every instance, the scopes of the statute's purposes are not fully specified.
4. As a result, the statements or working formulations of the statute's purposes require lexical refinements.

In addition, every statutory rule system emerges incomplete in most or all of the following twelve types of ways:

1. The rule system has gaps resulting in the need for additional positive, substantive rules, additions that, given the statute's purposes, ideally would have been included but were not, due (or on the evidence more likely than not due) only to oversight or failure to anticipate the need.
2. The rule system has gaps resulting in the need for remedial additions to the rule system, additions that for the same reasons should and would have been included.
3. The rule system requires exceptions to substantive rules that, given the statute's purposes and the factors that virtually universally limit the scope of legal rules, should have been provided and most likely, on the evidence, would have been provided, but were not due to oversight or the failure to antici-

¹²⁹ A statute can be quite complete in the sense that it covers the great majority of real world occurrences but, nonetheless, fails to address an unlimited number of less frequently or rarely occurring matters.

pate the need.

4. The rule system requires exceptions to remedial rules that, for the same reasons, should and would have been included.

5. The rule system requires exceptions to exceptions that, for the same reasons, should and would have been provided.

6. The rule system may be missing one or more subsidiary rule schemes necessary for implementation of primary rule schemes.

In addition, as noted above, in some instances, nothing in the text of the statute resolves the questions that ensue as a result of:

7. Internal conflicts within a statute;

8. Conflicts or apparent conflicts between the statute at issue and the Constitution, other statutes in the United States Code, or state law, with the latter possibly being pre-empted under the Supremacy Clause, and conflicts with other purposes operating in the law that may not appear in any statute;¹³⁰

9. Non-contemplated uses of a single statutory rule that produce seemingly unwanted results;¹³¹

10. Non-contemplated simultaneous uses of multiple statutory rules that produce seemingly untoward results;

11. Non-contemplated sequential applications of two or more

¹³⁰ The author believes that the avoidance of irrational, unfair, and unjust results in situations the legislature did not contemplate, especially when the purposes of a statute are not served or sufficiently served through its application in these situations, is an aim or purpose of the legislature, indeed, of society, even though there is no statute to that effect. We do not have statutes embodying every value that the legislature and society expect to operate in the law.

¹³¹ One of the more amusing of such uses may be apocryphal. The story is that during World War II a woman residing in a Southern state wanted to divorce her husband who was in Europe serving in the United States Army. Her attorney found a Civil War statute that had not been repealed which provided that any woman resident in the state could secure a divorce if it were proved that her husband was serving in the army of the United States of America. Divorce was granted on the basis of the statute, and the legislature hastily met to repeal it. Clearly such a statute was intended to be effective only at such time as there was a "war between the states." Not apocryphal at all is the use of a merger to affect an amendment of the articles of incorporation that is blocked by the articles of incorporation of the corporation that will disappear.

rules that produce seemingly untoward results; and

12. Failures to make it clear whether the statute entitles parties to private causes of action.

To reiterate, primarily for the benefit of parties of textualist leanings, on the vast majority of occasions, the statutory failings above are not only unintended but unwanted. They are not consciously intended limitations. They are not the result of political consensus and conscious exclusions from the law, but are almost always attributable to oversight and the inability to anticipate potentially infinite variations in fact situations. As is explained below, when this is so, lack of statutory text does not per se evidence an intention that such matters remain unaddressed. While the legislature may, and almost certainly does, mean exactly what it says with respect to matters well within its conscious contemplation, we cannot entertain a presumption to that effect with respect to matters overlooked or not anticipated.¹³² And there are thousands of such matters that ultimately surface in the wake of any statute that addresses even a moderately complex subject matter. The legislature, and society for that matter, must be viewed as intending that courts, the only parties in position to do so rationally, address these matters and address them in quite specific ways. Constitutionally, of course, if the evidence supports the proposition that, for example, what appear to be substantive and remedial gaps and omissions were intended,

¹³² Obviously, the following much cited statement needs to be modified: "Congress 'says in a statute what it means and means in a statute what it says there.'" *Hartford Underwriters Ins. Co. v. Union Planters Bank*, 530 U.S. 1, 6 (2000) (quoting *Conn. Nat'l Bank v. Germain*, 503 U.S. 249, 254 (1992)). When a party selects a particular dish from a specific, limited restaurant menu, one cannot conclude, on that basis alone, that the party would not have preferred a different dish that did not appear on the menu. Knowledge, external to the text of the menu, of the party's gustatory preferences might well permit one to assert with confidence that had Beef Wellington appeared on the menu, the party would doubtless have chosen it over all alternatives. Likewise, when a matter is not dealt with in a statute because it was overlooked or not anticipated, that does not mean that it would not have been dealt with had it been recognized. Further, the fact that it was not dealt with under such circumstances does not mean that one cannot say how it would have been dealt with, or likely would have been dealt with, had it been recognized.

or more likely than not intended, by the lawmakers, then there is, of course, a binding political decision that should be honored if it is constitutionally acceptable. But, in such instances, statutory purposes, as ends sought, have for various reasons been limited through the political process rather than through mere happenstance.

PART IV. THE INHERENT DEFICIENCIES IN THE COMMON,
ORDINARY, OR USUAL MEANINGS OF KEY STATUTORY WORDS
AND PHRASES

Due to recent advances in lexical semantics,¹³³ we are now in a position to understand the numerous problems that we encounter in the law with respect to words and phrases. Although an ambiguous ambiguity is almost universally cited as the *sole cause* of interpretive problems with respect to statutes, it should be obvious from the discussion above that neither lexical nor syntactical ambiguity is the cause of any of the sixteen types of problems identified in Part III. And, although each of the sixteen types of problems identified in Part III can give rise to ambiguity viewed as a conflict between, on the one hand, the purposes of statutory rule systems and other purposes simultaneously operative in the law and, on the other hand, the results of the application of the presumptive semantic content of statutory rules, i.e., an ambiguity of purpose, such conflicts are the result, not the cause, of the problems we encounter. Of equal importance, ambiguity is not even the primary lexical source of interpretive problems.

The meanings of words and phrases with which we are most often concerned in the law may be viewed in terms of both their extensions and intensions. The extension of a word or phrase that denotes a category we associate with either a noun or a verb consists of all instances of the denotata of the word or phrase. Thus, the extension of "cup" is all cups. The extension of "touch," as a verb, is all states and acts of touch-

¹³³ See, e.g., AITCHISON, *supra* note 21 and accompanying text; WIERZBICKA, *supra* note 23 and accompanying text; GEORGE LAKOFF, WOMEN, FIRE, AND DANGEROUS THINGS: WHAT CATEGORIES REVEAL ABOUT THE MIND (1987).

ing. A word can be defined ostensively through examples of its extension. That is, one way to define "cup" is "objects like this one here and that one over there," where the words spoken are tied to specific objects in the environment, each being a cup, by the gesture of pointing. One way to define "touch" is "states or acts like this one here and that one there," where the words spoken are tied to specific states or acts in the environment, each being a state or act of touching, by the gesture of pointing.

The intension of a word is a theoretically sound definition that is adequate to describe the category, or extension, the word is used to denote. As Anna Wierzbicka, one of the world's leading experts in lexical semantics, and others make clear, dictionary definitions are seldom intensions because dictionary definitions seldom are theoretically adequate.¹³⁴ For example, the primary definition of "cup" in the Webster's Third New International Dictionary of the English Language Unabridged (1986), i.e., that relating to a type of drinking vessel,¹³⁵ is approximately thirty five words. In contrast, Anna Wierzbicka's definition for "cup" requires two pages of text and addresses, at two levels, purpose, material, appearance, size, and use.¹³⁶ A definition for "touch" is not quite so lengthy. The non-metaphorical intension for "touch," as a non-active verb (as in "X is touching Y"), is something like: "to be in a spatial relation with another physical object whereby the subject is contiguous to and in direct contact with the other physical object." The non-metaphorical intension for "touch," as a transitive verb (as in "X is in the process of touching Y"), is something like: "an act whereby one physical object is brought into direct contact with another." Statutes involve encounters with at least twelve problems with the extensions and intensions—or meanings—of words and phrases. They also regularly involve a thirteenth set of problems with the

¹³⁴ See *supra* notes 21 and 24.

¹³⁵ There are also cups on golf courses into which golfers attempt to place golf balls.

¹³⁶ ANNA WIERZBICKA, *LEXICOGRAPHY AND CONCEPTUAL ANALYSIS* 26-27 (1985).

quantifiers, whether express or implied, "all" and "none."

Words and phrases used in a statute are often drawn from every day speech, other non-legal settings (including technical settings), or other legal settings in which linguistic categories have been established in light of purposes significantly different from those giving rise to the statute in which they appear. Since the scientific discipline of lexical semantics does not recognize the concept of "plain meanings," we shall continue to call these meanings "common meanings." Not surprisingly, the common meanings of key words and phrases used in a statute, being drawn from non-legal or significantly different legal settings, generally do not encompass categories of behaviors/actions, mental states, events, circumstances, situations, and the like, and combinations of the like, that are perfectly suited to the purposes of the statute into which they are introduced.¹³⁷ Time and again we discover, after the fact,

¹³⁷ There is a tendency to believe that we have lexically and conceptually divided and categorized the phenomena of the world in all possible ways. That is erroneous. The linguistic categories associated with words and phrases of the English language do not even begin to divide and classify the phenomena of the world into all possible categories. That would require an infinite vocabulary, a matter far beyond the capacity of a finite mind. Rather, as Derek Bickerton notes: "[W]e [lexically] distinguish only what we (as a species or as a cultural group sharing common values) need to interact with in significant ways." DEREK BICKERTON, *LANGUAGE AND SPECIES* 35 (1990). For example: "It has often been noted that no language has words with meanings like 'a left leg and left arm' or 'every other Friday'." *Id.* at 44. Of course, if there were a sport that regularly caused simultaneous injuries to the left arm and left leg, we might develop a special term.

As Alvin Goldman observes:

The creation of categories, kinds, or 'versions' is an activity of the mind or language. The world itself does not come precategorized, presorted, or presliced. Rather, it is the mind's 'noetic' activity, or the establishment of linguistic convention, that produces categories and categorical systems.

Goldman suggests viewing the relation between language and the world in terms not of a correspondence between words and pre-existing, external "facts," but in terms of "*fittingness*: the sense in which clothes fit a body."

The chief advantage of this metaphor is its possession of an ingredient analogous to the categorizing and statement-creating activity of the cognizer-speaker. At the same time, it captures the basic realist intuition that what makes a proposition, or statement, true is the way the world is.

that the classifications we need to fulfill the purposes of a given statute are generally not perfectly identical to the categories that satisfy everyday or non-legal needs or the needs of other statutes. The common meanings of the words and phrases that represent these categories have been established by societal usage conventions or legal concerns in some respects removed from the concerns of the statute in question. There is, accordingly, no reason to expect the categories, or extensions, these words and phrases have historically denoted in non-legal contexts or even in other statutes, and, therefore, their meanings, to be perfectly suited to the purposes of a given statute. Such problems are quite common in science. The category denoted by the every day folk biological term "animal" does not suffice for scientific biology. The concept of "force" in every day discourse does not suffice for physics.¹³⁸ Similar prob-

There are indefinitely many sorts of apparel that might be designed for the human body, just as there are indefinitely many categories, principles of classification, and propositional forms that might be used to describe the world. Although the body has parts, it is not presorted into units that must each be covered by a distinct garment. It is up to human custom and sartorial inventiveness to decide not only what parts to cover, but what types of garments should cover which expanses of body, and whether the garments should be snug or loose. . . .

Despite all this variety—humanly invented variety—there is still the question, for any specified type of apparel, whether a specific token of that type fits a particular customer's body. This question of fittingness is not just a question of style of garment. It depends specifically on *that* customer's body. Similarly, although the forms of mental and linguistic representation are human products, not products of the world per se, whether any given sentence, thought sign, or proposition is true depends on something extra-human, namely, the actual world itself.

GOLDMAN, *supra* note 11, at 152-53.

The fact is that we create the categories we find useful, when, as, and to the extent the need arises and is recognized. We find different categories useful at different times and for different purposes and create new categories, often by appropriating and adapting existing categories by metaphor, as new needs arise. See generally LAKOFF, *supra* note 133; GEORGE LAKOFF & MARK JOHNSON, *METAPHORS WE LIVE BY* (1980). That is, we linguistically and conceptually categorize and re-categorize the phenomena of the world as we find it useful to do so. There may be an infinite number of ways external reality could be divided and categorized, but we find only a finite number of such ways useful or interesting and most, therefore, are ignored. When we seek categories to suit our legal purposes, our legal purposes may not entail categories that have perfectly "fitted" other needs.

¹³⁸ When was the last time that it really hit you that force equals mass times

lems are encountered in the law.

The difficulty is that the ways in which the historical meanings of words and phrases are partially inappropriate to a statute's purposes, like the incompleteness of legal rules and rule systems, are generally not self-identifying or not fully self-identifying. It is generally impossible, *ex ante*, to recognize or ascertain the full, potential congruence or lack of congruence between the historical lexical categories and the lexical categories appropriate to a given statute in light of its purposes, although it is easy to do so *ex post* as the overlooked and unanticipated present themselves.¹³⁹ The following are twelve types of ways in which the meanings of nouns and verbs, and noun and verb phrases, from settings other than the statute at issue can be inadequate in light of the purposes of a given law.

1. To begin with the all too obvious, words and phrases used in statutes can be conceptually ambiguous. The primary form of ambiguity arises when a word denotes two or more relatively distinct extensions that are at least in part conceptually distinct. Although this form of ambiguity is often treated as *the* essential source of the need for statutory interpretation, it is, in fact, the least of the lexical problems in the law because context resolves most ambiguities so effectively that they are never even noticed.¹⁴⁰ As noted, the term is all too often used when any problem relating to statutes arises.

2. Words and phrases used in statutes can evidence non-conceptual ambiguity due to prototypes, or prototypical extensions, and non-typical category members, or non-typical exten-

acceleration?

¹³⁹ Item 6, *infra* p. 118, and Item 8, *infra* p. 122, fully illustrate this point.

¹⁴⁰ For example, if one says of a known alcoholic, "He's drinking again," even though there are many things a person can drink and the elided direct object is, accordingly ambiguous, almost no one will ever notice the ambiguity. Rather, hearers will automatically understand "alcoholic beverages," and the ambiguity will pass unnoticed. The example derives from LEVINSON, *supra* note 30, at 176. In a torts class, the statement "The defendant was responsible for the battery" would almost certainly be immediately understood to mean that the defendant was liable for the tort of battery. It would occur to few that the battery in question might be a flashlight battery or an assemblage of artillery.

sions.¹⁴¹ Prototypes result when a category has one extension comprised of typical members that are generally the subject of reference in a linguistic community but also conceptually includes an extension of non-typical members that are not generally the subject of reference.¹⁴² That is, there is a prototypical extension and a non-typical extension, even though both extensions are covered by the same general concept. The existence of prototypes gives rise to a form of ambiguity that is more subtle than conceptual ambiguity. For example, in America, the typical bird is a robin or sparrow. Thus, if someone says that he has a number of birds on his lawn, one will assume that the party has typical birds on his lawn, not a flock of ostriches or condors, even though, conceptually, ostriches and condors are unquestionably birds. The typical is viewed as intended in use—i.e., the default—unless further clues are provided. In the law, issues deriving from prototypes occasionally surface.¹⁴³

¹⁴¹ We can add non-lexical, syntactical ambiguity to the list of problems encountered in the drafting of statutes. For a thorough discussion of the problem, see SOLAN, *supra* note 16, at 28-92.

¹⁴² On prototypes, see, for example, LAKOFF, *supra* note 133.

¹⁴³ *Smith v. United States*, 508 U.S. 223, 225 (1993) addressed the meaning of the phrase "use a firearm" in 18 U.S.C. § 924(c)(1) (2000), which enhances the prison term for drug crimes when a party uses a firearm in connection with a drug crime. The phrase has both prototypical and non-typical meanings. In *Smith*, a party, the purchaser of drugs, attempted to use an automatic weapon as payment for drugs, and the government sought to enhance the penalty on the grounds that the party "used a firearm" in connection with the crime. *Id.* at 226-27. The majority of the justices on the Supreme Court agreed that the defendant had used a firearm within the meaning of the statute. *Id.* at 241. Justice Scalia dissented on the grounds that the phrase typically is used to mean such things as to threaten or to inflict injury with a gun. *Id.* at 242. Justice Scalia was certainly correct in one sense. Although on abstract conceptual grounds, the phrase can include use of a firearm as a hammer, door-stop, medium of exchange, or fishing line sinker, those are not the prototypical uses. Of course, the mere fact that the court majority opted for a non-typical meaning is hardly proof of error. See also *In re Pizzi*, 153 B.R. 357, 362 (Bankr. Fla. 1993) (holding that what is conceptually clearly an annuity, although not a typical annuity, is not an annuity); *Int'l Bhd. of Teamsters v. Daniel*, 439 U.S. 551, 566-67 (1979) (holding, on the basis of several patently flawed economic and financial arguments, that what is clearly conceptually an investment contract under *SEC v. W. J. Howey*, 328 U.S. 293 (1946), albeit not a typical investment contract, is not an investment

3. The common meanings of words and phrases used in statutes can be vague in varying degrees. Vagueness is distinct from ambiguity. Ambiguity exists when there are two (or more) distinct meanings (extensions and corresponding intensions) in actual use and it is not clear from context which is intended. Vagueness exists when and to the extent a word fails to denote with certainty.¹⁴⁴ Vagueness occurs when and to the extent a word's extension is unsettled because there is no established usage convention in a linguistic community over a portion of its potential range of use.¹⁴⁵ To the extent

contract).

¹⁴⁴ Creating a phrase often fails to eliminate vagueness. The phrase "major life activity" used in the Americans with Disabilities Act, 42 U.S.C. §§ 12101-12117 (2000) (relating to employment), is not rendered precise by inclusion of the word "major." That word, itself, is vague. It is what Anna Wierzbicka calls an "approximative." Anna Wierzbicka, *Precision in Vagueness: The Semantics of English Approximatives*, 10 J. OF PRAGMATICS 597 (1986). As for "major life activities," we have no exhaustive list of major life activities that is conventionalized in the lexicon.

¹⁴⁵ That is, the word's or phrase's extension, or the category members it denotes, is not denoted precisely and does not constitute a conventionally established, exhaustive, and discrete set, i.e., the set is potentially open—although some members of the set are unequivocally designated by convention. In terms Paul Ziff employs, there is no established regularity of usage with respect to some possible members of the extension. ZIFF, *supra* note 22. For example, Ray Jackendoff examines the meaning of the verb "defrost" and finds that that there are instances of behavior that are clearly instances of defrosting, instances in which one is unsure, and other instances where one is clear that the behavior cannot be called defrosting. RAY JACKENDOFF, *PATTERNS IN THE MIND* 198-99 (1994). He observes:

The point is that everyone agrees on their judgment of the ideal situation. As we move away from the ideal, there is a broad gray area or penumbra of cases that "sort of" fall under the concept, and about which people may disagree somewhat. Then, as we move still further away from the ideal, everyone is in agreement again that the situation does not fall under the concept.

Id. at 199. Professor Jackendoff then points out that "[i]n the case of the concept of defrosting, the existence of a gray area is harmless enough. Why bother worrying about it?" *Id.* That provides the explanation for the gray areas. There are no needs for societal conventions in the gray areas, at the peripheries of the uses of words, because the need for a convention at the peripheries arises too rarely either to require or to occasion a societal convention. Were there needs for societal conventions at the peripheries, i.e., greater specificity or precision, they would exist. Nonce determinations, whether arbitrary or not, can generally be made to take

there is no established convention of inclusion in or exclusion from a word's or phrase's extension, one obviously cannot establish by convention what the word denotes and, therefore, cannot say what it means, i.e., stands for or represents.¹⁴⁶

Almost all words that denote become vague at the peripheries of their usage. Some are vague at their cores. As a practical matter, vagueness results from the fact that in many practical and communicative settings we do not need extremely fine denotational, or extensional, precision and, being economical of our time and efforts, we, accordingly, do not bother to establish denotational, or extensional, conventions with precision beyond our needs.¹⁴⁷ In the non-legal settings in

care of the rare occasions when a convention is required.

For example, there is no word or phrase in common usage that captures with precision what the phrase "major life activity" in the Americans with Disabilities Act, 42 U.S.C. §12102(2) (2000), purports to denote. Being precise in that specific way is not a matter of societal concern outside the law. As a result, merely slapping together the words "major," "life," and "activity" does not create the precise category needed in the law because the concept constituted by the phrase "major life activity" is not conventionalized in the American-English lexicon. In particular, the term "major" is an approximative. See Wierzbicka, *supra* note 144. The meaning of an approximative, qua definition, may, as Anna Wierzbicka has demonstrated, be precise enough, see *id.*, but the application of the term, as is true of all vague terms, cannot be precise in all instances. This, of course, means that fact finders asked to apply such words are being asked to "find" conventions when none exist; that is, to establish a convention where none existed and to make law. Further, different fact finders can make contradictory law on identical facts. In the case of the Americans with Disabilities Act, not only is there no word or phrase in the American-English lexicon that provides the precise conceptual distinctions for the category the statute needs, the parties drafting the statute were obviously not themselves clear regarding which "life activities" were to be considered as within the scope of the statute's concern, much less all that were to be considered "major." Since they did not know in all instances what they intended, they chose a phrase that is clearly vague at the margins.

¹⁴⁶ With words that have prototypical and non-typical meanings, one encounters both ambiguity and vagueness. At some point even the conventions regarding the non-typical run out. Is a cliff dwelling that was used by primitive people, and now is used occasionally by campers, a dwelling for purposes of burglary? There is simply no preexisting convention that says one way or the other.

¹⁴⁷ When the available, historical conventions do not provide the needed precision, we create the precision through special definitions. That is the reason that we create technical definitions for special fields of study. Previously, the precise definition that constitutes a technical definition was not needed. When the need is encountered and prior definitions are recognized as inadequate, the technical

which the meanings of words and phrases evolve and are established through usage conventions, the lack of thoroughness or precision in conventional usage poses no problems for linguistic communicators. The words, though vague, suit our needs. When, however, by statute we create entitlements, rights, and obligations that turn on the classifications in question, there finally emerges for the first time in human history a need and a reason for persons to seek thoroughness and precision. Someone stands to win or lose a damages award, tax deduction, or contingent fee. Unfortunately, in many instances there is no pre-existing convention, no thoroughness, and no precision, and no amount of judicial linguistic analysis or resort to canons of construction will uncover any.¹⁴⁸

4. The common meanings of words and phrases used in statutes can exhibit vagueness in the form of binary contrasts. When the concepts associated with two words putatively split phenomena into two mutually exclusive groups but, due to vagueness, i.e., the lack of a convention that covers all possible combinations of "facts," fail to do so, one is often faced with phenomena, often behaviors and relations between persons and institutions, that do not clearly fall under either concept. Thus, we encounter such traditionally contrasting poles such as: "hot" and "cold," "large" and "small," and "tall" and "short."¹⁴⁹ In the law we encounter poles such as "dwelling" and "non-dwelling," "material" and "immaterial," "substantial" and "insubstantial," "employee" and "independent contractor," and "debt" and "equity." The extremes of the polar contrasts are clearly enough established by convention, but

definition is created.

¹⁴⁸ Anyone who has had to classify jobs as "management," "administrative," and the like under the Fair Labor Standards Act Regulations (29 C.F.R. §§ 541.1-.2, 541.102-.115, 541.200-.210 (2005)) has encountered vast gray areas. There are no conventions that dictate precisely for each business, regardless of its size, where jobs begin and cease to be management or administrative jobs. What may be viewed as a management function in a small business may be viewed as ministerial in a large business.

¹⁴⁹ The meanings of these words are also highly context sensitive. "Large" in "large elephant" obviously signifies something different than "large" in "large microbe."

the middle ground between them has not by convention been classed under either pole. There has been no reason to be-
cause nothing particularly useful or important to anyone has
ever turned on precisely classifying the middle between the
extremes. But all that changes when the law enters the pic-
ture. The law requires precise, exhaustive classification, but
there is none.¹⁵⁰

5. The common meanings of words and phrases used in stat-
utes can evidence a form of vagueness that gives rise to the
difference between rules and standards. A rule is a specific
and focused convention. A standard is an unavoidably vague
and diffuse convention that purports to be complete and ex-
haustive but, in fact, has substantial gaps not covered by
convention. The extensions of the key words and phrases in a
rule are very specific. In the case of rules, we make sure there
is little vagueness. For example, most states mandate that
when a school bus driver approaches rail road tracks, the
driver is to stop the bus before reaching the tracks, open the
door, look both ways, ensure that no train is approaching, and
close the doors before proceeding. Thus, the actions that con-
stitute "due care" in that setting are precisely specified, i.e.,
defined. We do not want to leave the matter to individual
judgment. Accordingly, the rule itself precisely prescribes the
conduct that satisfies it.

Standards arise from a need to distinguish between per-
missible and non-permissible behaviors in a vast number of
arenas of action or activity regarding which there are in many
instances no precise conventions as to what is proper or not,
and, equally important, cannot be because of the vastness of
the number of potential fact situations. That is, no one has the
time or ability *ex ante* to identify each situation in which the
concerns underlying the standard might prevail and to pre-

¹⁵⁰ What investor, *qua* investor, cares whether an instrument that has both
debt and equity characteristics, such as a "bond" that pays both fixed interest
and additional interest measured as a percentage of profits, or preferred stock
with a mandatory fixed rate of return, is "really" debt or equity—it simply is
what it is—unless, for example an income deduction turns on it.

scribe a rule of conduct specifically for each such situation. We resort to standards, and rationally so, in an attempt to accomplish what otherwise could not be accomplished. Thus, in torts we use a standard of due care, even though we have no precise societal conventions that categorize all potentially harmful behaviors in all conceivable circumstances as the exercise *vel non* of due care. Malpractice, in contrast, establishes precise rules based on professional conventions where standards would otherwise prevail.¹⁵¹

6. The common meanings of words and phrases used in statutes can conceptually misclassify. In doing so, they can include or exclude facts that *ex post* should clearly or clearly not be within their extensions in light of the purposes of a given statute. The problem of conceptual misclassification can be illustrated by the term "public offering" in the transaction exemption (itself, literally, a securities exemption) of subsection 2 of § 4 of the Securities Act of 1933.¹⁵² The traditional meaning of "public offering" was conceptually inappropriate, *i.e.*, misclassified, in that it would have in some instances denied the act's protection to parties clearly in need of same. The Supreme Court provided a conceptually distinct and appropriate definition that correctly classified by looking specifically to the need for the statute's protections.¹⁵³ The defi-

¹⁵¹ Although in one sense all vagueness is vagueness, that is the absence of a convention regarding the extension of a word or phrase that clearly places certain phenomena within or without the extension of the term at issue, vagueness creates several distinct types of logical and practical problems. For example, the vagueness in 3, above, prevents parties from saying unequivocally that a given phenomena is either A or not-A. The vagueness in 4, above, prevents parties from saying unequivocally that a given phenomenon is an A or a B. For the problems those types of vagueness entail for philosophy, especially logic, see TIMOTHY WILIAMSON, *VAGUENESS* (1996). The type of vagueness in 5, above, in effect requires juries to make law, that is to decide when a party has or has not breached a duty when, but only when, there is absolutely no pre-existing societal convention (remember, reasonable persons can disagree!) regarding the propriety of the conduct at issue. That is, we ask juries to establish particular, case specific, after the fact, conventions when, but only when, there is no pre-existing convention to guide the conduct of the party charged with wrong doing! Does that really comport with due process or principles of fundamental fairness?

¹⁵² 15 U.S.C. §§ 77a-aa, 77d(2) (2000).

¹⁵³ *SEC v. Ralston Purina Co.*, 346 U.S. 119, 125 (1953). Subsection 2 of sec-

tion 4 of the Securities Act of 1933, 15 U.S.C. § 77(d) (2000), exempts from the registration requirements of the Act "transactions by an issuer not involving a public offering." The term "public offering" was not a technical, legal term, but rather one deriving from the securities industry. The securities industry uses the term "public offering" to designate and mean something like "securities offerings made broadly through the underwriting process, with the capital of all comers sought and welcome." A non-public, or private, offering could, accordingly, be defined as "any offering of securities that the offering group limits to a few individuals or institutions that they approve or a class of individuals or institutions that they approve." The securities industry's classification of "public offering" was not, however, fully appropriate to the purposes of the Securities Act of 1933. For the purposes of the Act an offering not involving a public offering, that is, a non-public offering, is not one wherein the offerors restrict the offering to a small group, but rather one in which the offering group makes its offer to those individuals and institutions that do not need the protections of the mandatory disclosure requirements of the act because they have access to the information that such requirements mandate be disclosed. There was no reason for the securities industry to classify in the same manner that the law classifies. The purposes were different. In particular, the securities industry was not concerned with protecting persons who needed the protections of the Act. If anything, the industry for years aimed at the efficient fleecing of such persons. Consider, for example, the market manipulations of Jay Gould and Commodore Vanderbilt. The industry's abuse of such persons was one of the very reasons for the enactment of the Securities Act of 1933. Therefore, there was no reason why the category the securities industry designated by the term "public offering" should be based on the need for the protections of the Act. This illustrates the fact that categories relate to purposes in classification, and differences in those purposes will lead to different categories.

The need to redefine the term "public" and its correlate, "non-public," became obvious once the misclassification was revealed by consideration of the statute's purposes. Given the purpose of protecting individuals who do not have access to the information securities registration will disclose, the public, for purposes of the Securities Act of 1933, is comprised of all persons who need the protections of the Act, i.e., mandatory disclosure of material information regarding the issuer; and, accordingly, a public offering under the Act is an offering of securities made to members of that class. As the Supreme Court, in a pragmatically (in the linguistic sense) correct, indeed, in a pragmatically mandated, opinion, held in *Ralston Purina* that a non-public offering is an offering made to parties who do not need the protections of the act, while a public offering is an offering made to parties who need the protections of the act.

Conceptually, the definitions of "public offering" and "non-public offering" as operative in the securities industry and the Act were radically different. The categories that each concept denoted or described, however, coincidentally substantially overlapped. The historical meaning of the term as originally appropriated to the Act was highly useful in that it covered a very large portion of the individuals who would need the protection of the Act, but, as *Ralston Purina* made clear, it did not cover all. *Id.* at 126. Radical conceptual reclassification was necessary

tion was also new to the lexicon. Indeed, the definition could not have been in the lexicon at any time prior to the date the President signed the Securities Act of 1933 into law.

7. The common meanings, viewed primarily as extensions and secondarily as intensions, of words and phrases used in statutes can be over-inclusive in light of a statute's purposes and overriding considerations. Possibly the most famous, or infamous, example is the phrase "relates to" used in the preemption clause of ERISA, a clause that has generated numerous interpretive decisions and a good many scholarly articles.¹⁵⁴ ERISA's preemption provision purports to preempt all state laws that "relate to" an employee benefit plan subject to ERISA. Everything, of course, relates to everything else in some way or ways. Some of the many ways that state laws can relate to employee benefit plans is by legally empowering them, subjecting them to legal obligations, and affecting their costs. Taken literally, ERISA's "relates to" clause would have, *inter alia*, eliminated malpractice claims against plan medical providers by participants in employer sponsored medical plans, freed employee plans subject to ERISA from contract obligations as well as shielded those who deal with plans from contractual obligations to the plans, and shielded service providers to plans, such as lawyers, accountants, and actuaries, from malpractice claims by plans. Applied literally, the preemption provision would have rendered plan operation impos-

to include a small, additional group needing the protections of the Act. The historically established meanings of "public offering" and "non-public offering," not only were initially partially misclassified for purposes of the statute, but also, the corrected, statutory meaning constituted a classification that was not previously encoded in the lexicon of the American-English language. Rather obviously, it was a classification that, in fact, could not have existed prior to the passage of the Securities Act of 1933.

¹⁵⁴ 29 U.S.C. § 1144 (2000), referred to by practitioners as ERISA § 514, provides *inter alia*:

(a) Except as provided in subsection (b) of this section, the provisions of [this title and title IV] shall supersede any and all State laws insofar as they may now or hereafter relate to any employee benefit plan described in section 4(a) and not exempt under section 4(b)

sible. Needless to say, courts have not allowed the reach of ERISA's "relates to" clause to reach so far and have narrowed its extension and corresponding intension.¹⁵⁵ The resulting

¹⁵⁵ "Relate to" may be rephrased (not defined as) "stand in some relation to." Phenomena can stand in, e.g., temporal, causal, affective, magnitudinal, directional, proximity, containment, genetic, familial, functional, and conceptual relations to other phenomena. Justice Souter has observed: "If 'relate to' were taken to extend to the furthest stretch of its indeterminacy, then for all practical purposes pre-emption would never run its course, for '[r]eally, universally, relations stop nowhere" N.Y. State Conference of Blue Cross & Blue Shield Plans v. Travelers Ins. Co., 514 U.S. 645, 655 (1995) (citation omitted). "[I]nfinite relations cannot be the measure of pre-emption." *Id.* at 656. As Judge Karen J. Williams has even more pointedly observed: "Taken at its face value, the term 'relates to' has no logical boundary. . . . In one way or another, everything relates to everything else." Metro. Life Ins. Co. v. Pettit, 164 F.3d 857, 861 (4th Cir. 1998). That is, everything stands in *some* relation to everything else. If in no other way, everything exists either before, after, or at the same time as everything else and relates to everything else by existing either before, after, or simultaneously with everything else. One could not employ a more comprehensive term. One significant way state laws can relate to employee benefit plans is by having an effect on their administration, their costs, and the like. In spite of the breadth of the phrase courts have over the years pared back, as not counting for purposes of ERISA's preemption provision, the number of ways that state laws can relate to plans subject to ERISA and be preempted. Thus, although state malpractice laws clearly affect employer sponsored health plan costs, ERISA does not, with respect to health plans, preempt state malpractice laws. It would hardly seem consonant with a law designed to protect employees to allow physicians providing services through employer sponsored health plans to commit malpractice at will, even though the reduction in risk exposure to physicians would likely lower plan costs. *See, e.g.,* Moreno v. Health Partners Health Plan, 4 F. Supp. 2d 888 (D. Ariz. 1998); Prihoda v. Shpritz, 914 F. Supp. 113 (D. Md. 1996); Chaghervand v. CareFirst, 909 F. Supp. 304 (D. Md. 1995); Haas v. Group Health Plan, Inc., 875 F. Supp. 544 (S.D. Ill. 1994); Independence HMO, Inc. v. Smith, 733 F. Supp. 983 (E.D. Pa. 1990). On the other hand, if a claim relates to a plan denial of medical coverage, even if based on a medical judgment, it is deemed related to plan administration and is pre-empted. *See, e.g.,* Brock v. Primedica, Inc., 904 F.2d 295 (5th Cir. 1990); Thompson v. GenCare Health Sys., 49 F. Supp. 2d 1145 (E.D. Mo. 1999), *aff'd* 202 F.3d 1072 (8th Cir. 2000); Clark v. Humana Kan. City, Inc., 975 F. Supp. 1283 (D. Kan. 1997). *Contra* Cicio v. Does, 2003 U.S. App. Lexis 2925 (2003).

In addition, as a practical necessity it was recognized that employee benefit plans must be allowed to engage in transactions in pursuit of their business and that such transactions must be subject to, e.g., state commercial laws, even though of course such state laws clearly relate to the employee benefit plans that engage in commercial transactions. *Smith v. Provident Bank*, 170 F.3d 609 (6th Cir. 1999). Accordingly, claims against employer sponsored employee medical plans

meaning, both extension and intension, is unique to the both the lexicon and the United States Code and is still developing. 8. The common meanings, viewed primarily as extensions and secondarily as intensions, of words and phrases used in statutes can be too restricted or under-inclusive. For example, the historical meaning of the word "discriminate" as used in Title VII of the Civil Rights Act of 1964 ("Title VII"),¹⁵⁶ the paramount purpose of which was to ensure equality of employment opportunity without regard to race, color, sex, religion, or national origin, proved to be too restricted. "Discriminate" is a term that traditionally denoted intentional conduct. To serve the purposes of Title VII, the meaning had to be expanded to include forms of unintentional conduct.¹⁵⁷

by third-party providers for negligently affirming insurance coverage for parties who in fact are not covered are not pre-empted. *Meadows v. Employers Health Ins.*, 47 F.3d 1006 (9th Cir. 1995). Likewise, claims by employers or trustees against consultants for negligence, malpractice, and common law fraud are not preempted. *Forbus v. Sears Roebuck & Co.*, 30 F.3d 1402 (11th Cir. 1994), *cert. denied* 513 U.S. 1113 (1995); *Airports Co., Inc. v. Custom Benefits Servs.*, 28 F.3d 1962 (10th Cir. 1994); *Carl Colteryahn Dairy, Inc. v. W. Pa. Teamsters & Employers Pension Fund*, 785 F. Supp. 536 (W.D. Pa. 1992). Because courts have rightly judged that a provision put into the law at the last minute without much forethought was not intended to lead to the preemption of numerous areas of state regulation just because in some remote way they relate to employee benefit plans subject to ERISA, they time and time again decide that given ways of relating to ERISA governed plans do not fall within the extension of "relate to" for purposes of ERISA's preemption clause. For extended discussions of ERISA's preemption clause, see, for example, Jeffrey A. Brauch, *ERISA at 25—and Its Most Persistent Problem*, 48 KAN. L. REV. 285 (2000); Howard Shapiro, René E. Thorne & Edward F. Harold, *ERISA Preemption: To Infinity ad Beyond and Back Again? A Historical Review of Supreme Court Jurisprudence*, 58 LA. L. REV. 997 (1998); Karen A. Jordan, *The Shifting Preemption Paradigm: Conceptual and Interpretive Issues*, 51 VAND. L. REV. 1149 (1998); Catherine L. Fisk, *The Last Article About the Language of ERISA Preemption? A Case Study in the Failure of Textualism*, 33 HARV. J. ON LEGIS. 35 (1996).

¹⁵⁶ 42 U.S.C. §§ 2000e to 2000e-17 (2000).

¹⁵⁷ Title VII contains a substantive rule that prohibits discrimination in employment based on the fact that an individual falls into any of its protected categories. *Id.* Until the enactment of Title VII, the term "discriminate" had always meant "to distinguish and sometimes also to treat differently on the basis of some consciously and intentionally applied standard or recognized differences in characteristics." One discriminates, for example, between the bouquets of a Burgundy and a Cote du Rhone, between true and false friends, between the highly talented

9. The common meanings of words and phrases used in statutes can denote excessively general categories, or extensions, given a statute's purposes. That is, the common conventions underlying the uses of words and phrases may not accommodate distinctions required in light of a statute's purposes. If the facts in question were not surfaced or anticipated at the time a statute was drafted, the fact that the categories associated with a statute's words or phrases fail to accommodate distinctions needed in light of the statute's purposes will have been noticed by no one. Accordingly, the lack of necessary distinctions will not have been addressed at that time.

To take an example from the criminal law, there are no words in common usage, ethics, or theology that reflect the distinctions required in criminal law relating to wrongful killing. The technical word "homicide" is too broad, as a killing in self-defense is a homicide although it is not a wrongful

and less talented in sports tryouts and musical auditions, and between the standard harmonic progressions of J.S. Bach and Gustav Mahler. An examination of the contexts of the uses of "discriminate" outside of Title VII indicates that the term had always denoted conscious and intentional conduct. In *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), the question was presented: can an employer be liable under Title VII for discriminating if it unintentionally excludes a protected class from employment. Specifically, can an employer be deemed to discriminate if, without an intention to discriminate, i.e., treat differently, it employs a facially neutral hiring criterion that is not related to the ability to perform the job in question but that has an adverse impact on the employment opportunities of a group protected by the Act? The Supreme Court held, in a pragmatically correct, indeed, in a pragmatically mandated opinion, that an employer can, indeed, be liable under such circumstances. *Id.* After the Court's decision in *Griggs*, it could be said that "discriminate" means "(a) to distinguish and sometimes also to treat differently on the basis of some consciously applied standard or recognized differences in characteristics and (b) under Title VII, in addition to the conduct in (a), unintentionally to treat differently by the use of a facially neutral criterion that is unrelated to the performance of the job if the result adversely affects the employment opportunities and other employment related rights of a class protected by Title VII," with the material in italics being both an expansion of the historical meaning and a new meaning introduced into the American-English lexicon. Accordingly, the historical meaning, the meaning to be found in the American-English lexicon prior the enactment of Title VII, was too restricted, or under-inclusive, in light of the statute's purposes. That fact was not apparent at the time the statute was enacted and did not surface until almost seven years after enactment.

killing. The phrase "wrongful killing" is also too broad, as it does not accommodate the multiple distinctions exhibited in first degree murder, second degree murder, voluntary manslaughter, involuntary manslaughter, negligent homicide, and the like — distinctions deemed appropriate in light of levels of punishment deemed appropriate in light of differences in mental state. Those categories had to be created in light of the needs of the law.¹⁵⁸

10. The common meanings of words and phrases used in statutes that capture paradigmatic facts can fail to encompass what were above called non-conceptual and substantial equivalents.¹⁵⁹ When that is so, the extensions, and, therefore, the intensions, of the words and phrases will be too restricted. As noted above, the parties who draft statutes may identify what they believe to be all the sets of elemental facts that, given the statute's purposes, should bring the statute's substantive and remedial schemes into play. But the reality is that it is often not possible *ex ante* to ensure that one has identified all such facts. Some such facts may not even exist until devised through the machinations of skilled lawyers reflecting in leisure after the frantic rush that led to the enactment of a statute. Because non-conceptual and substantial equivalents are often extremely difficult to uncover, the legislature cannot take the time to dream up all the possible non-conceptual and substantial equivalents, and accordingly they are often overlooked, not anticipated, and not lexically provided for in a legal rule's coverage.¹⁶⁰

¹⁵⁸ In the statutory arena, the term "income" proved too general for the tax laws which have found it necessary to distinguish ordinary income from capital gains.

¹⁵⁹ See *supra* at pp. 95-97.

¹⁶⁰ Any time the legislature expressly or by implication purports to establish an exclusive list of conduct that is, e.g., prohibited because it leads to an unwanted end, one can be certain that parties who want to achieve that end will attempt to concoct non-conceptual or substantial equivalents to the proscribed conduct in an attempt to evade the strictures of the law. Should a statute read: "Thou shalt not beat thy neighbor with stick or stone," one can be assured it would not be long before crow-bars would become popular items with the rough crowd. That is an obvious case, but often functional equivalents are often concoct-

11. The common meanings, viewed primarily as extensions and secondarily as intensions, of words and phrases used in statutes can fail to exclude exceptional fact aggregates. As noted above, facts that have been identified and captured by words or phrases used in a legal rule can occur in conjunction with additional facts that were not initially contemplated, and it may not be proper, given the purposes of the rule or overriding considerations,¹⁶¹ to include within the rule's coverage the aggregate of the facts identified and the additional facts, i.e., exceptional fact aggregates. To the extent exceptional fact aggregates are not identified *ex ante*, they will not be expressly excluded from a legal rule's coverage by exclusion from the extensions of the rule's words and phrases.

12. The common meanings, viewed primarily as extensions and secondarily as intensions, of words and phrases that denote exceptional fact aggregates may not exclude extra-exceptional fact aggregates that constitute exceptions to exceptions that reinstate the application of a rule.¹⁶²

13. The thirteenth problem with words arises whenever a

ed by counsel in leisure during the months and years following the enactment of a statute.

¹⁶¹ This situation is aptly illustrated with the common law concept of battery, which was pared back in the case of conduct that would otherwise constitute battery to take into account additional facts that have given rise to privileges such as self-defense, law enforcement, and consent.

In the statutory arena, the Sherman Act, 15 U.S.C. §§ 1-7 (2000), textually fails to exclude from its strictures numerous exceptional fact aggregates, specifically those known as "reasonable restraints of trade." Rather, the act literally prohibits all restraints of trade, and the Supreme Court originally so interpreted it. See *United States v. Trans-Mo. Freight Ass'n*, 166 U.S. 290 (1897). It was not long before the courts began to recognize that reasonable restraints should be excepted. See *United States v. Addyston Pipe & Steel Co.*, 85 F. 271 (1898).

¹⁶² To again lean on the common law, the privilege of self-defense in battery, an exceptional fact aggregate of a battery plus justifying circumstances, is subject to the requirement that one exercising the privilege not exceed reasonable force. By analogy, the Americans with Disabilities Act of 1990, 42 U.S.C. §§ 12101-12213 (2000), permits an employer to refuse to employ a disabled person who cannot perform the essential functions of the job. However, an employer that asserts that a disabled individual cannot perform the essential function of the job can, nonetheless, be held liable if the disabled person could perform the job if provided a reasonable accommodation.

statute expressly or implicitly covers all conduct, occurrences, situations, and the like, or excludes from its coverage all conduct, occurrences, situations, and the like. Whenever there are overlooked or unanticipated exceptions, rather obviously the word "all," or the implicit quantifier "all," means "some." Thus, although § 514(a) of ERISA provides that the provisions of titles II and IV of ERISA "shall supersede *any* and *all* State laws insofar as they may now or hereafter relate to any employee benefit plan,"¹⁶³ courts have held on many occasions that given state laws that clearly relate to employee benefit plans are not preempted by ERISA.¹⁶⁴ Courts have determined that full scale preemption would be absurd. Accordingly, the precise and unambiguous terms, "any" and "all," terms that by canon supposedly require no interpretation, as used in § 514(a) of ERISA, in fact, mean a quite indefinite "some." So it is with respect to many statutes.

Due to the thirteen factors, above, the presumptive lexical meanings of words and phrases used in statutes may not be what is fully appropriate or required in light of the statute's purposes and other purposes operating in the law. The meanings of the words and phrases used generally clearly cover the paradigmatic facts a statute is designed to address. Those facts are recognized *ex ante*. If the common, usual, or standard meanings of words and phrases do not capture the paradigmatic facts, that fact will be recognized and the words and phrases will be supplied with special definitions sufficient to capture them. But the drafters of statutes cannot generally *ex ante* recognize these lexical inadequacies or cannot recognize all of them, and, therefore, cannot address them or all of them *ex ante*. Therefore, courts are often forced *ex post* to redefine words—*not simply resolve lexical ambiguities*—to bring their meanings into line with the purposes of a given statute, often creating meanings that are new to the human lexicon.¹⁶⁵

¹⁶³ 29 U.S.C. § 1144(a) (2000) (emphasis added).

¹⁶⁴ See *supra* note 155.

¹⁶⁵ It may be worthwhile to note that some matters can be addressed by either

CONCLUSION

With the fundamental problems that we face in formulating statutory rule systems identified in significant detail, it is clear that the fundamental objective in statutory interpretation must be to accomplish *ex post* what has unintentionally and unavoidably been left undone in the formulation process. That is the only way to avoid unintended, random, accidental, arbitrary, and often irrational, unfair, or unjust results when overlooked and unanticipated situations surface—results that, even with the utmost diligence, often could not *ex ante* be avoided by express statutory provisions or express definitions and that run counter to the purposes of the statute at issue and other purposes simultaneously operating in the law. Avoiding such results requires: (1) ascertaining the purposes underlying statutory rule systems and, when, and to the extent that, the purposes are inchoate or unrefined, properly clarifying or refining them, (2) using the purposes underlying statutory rule systems and other purposes simultaneously in operation in the law to complete inevitably incomplete statutory rule systems in consonance with those purposes when and as often as any of the twelve types of incompleteness identified in Part III surface, and (3) using the purposes in (2) to modify the meanings of words and phrases used in statutory rule systems to bring those meanings into line with the purposes when any of the thirteen types of deficiencies in lexical

the addition of rules to a statute or by the equivalent modification of the meanings of words and phrases. These alternatives arise whenever it is appropriate either to expand or contract a category of covered, or elemental, facts. Such expansions or contractions can be accomplished either by special rules expressly including or excluding appropriate category members or by redefining words and phrases to include or exclude the appropriate category members. Thus, if it is appropriate, in connection with a given labor or employment law, to include independent contractors among the protected group along with common law employees, the courts can either hold (a) that in addition to common law employees, an expressly protected category, the purposes of the law demand that independent contractors also be covered or (b) that for purposes of the statute in question the word "employee" includes independent contractors. The substantive result is the same. Courts often appear to prefer covertly to redefine to avoid the appearance "making" law.

meaning identified in Part IV surface. Accomplishing those three tasks simultaneously resolves ambiguities.

Ideally, we would like a process that enables us to accomplish the three tasks in a manner that (1) is fully consonant with the constitutional principle of legislative supremacy and that is, therefore, entirely objective and politically and ideologically neutral; (2) yields predictable results in all cases; (3) is epistemologically unassailable; and (4) operates in accordance with the cognitive processes whereby we comprehend language—which processes, not surprisingly, generally operate in accordance with the principles of epistemology. While identifying the problems we encounter in formulating statutory rule systems reveals the process for addressing them, we cannot achieve the ideal. Nonetheless, we can, fortunately, approach the ideal.

Reflection reveals that the problems we encounter with respect to statutes are common in human experience, and the process for addressing them is readily identifiable and amazingly uniform and constant across a broad array of human activities. The adaptations of this process to accommodate the occasional special needs of the law are minor and obvious. In general terms, formulating statutes requires us to establish schemas, plans or designs if you will, for accomplishing given purposes in the face of eventualities—a number of which inevitably have been overlooked and others of which have not been, in many instances, could not be, anticipated. That is common in human experience. There are thousands of other settings in which we face the identical problems, in which we must formulate plans or designs for accomplishing given purposes in the face of overlooked and unanticipated occurrences or events.¹⁶⁶ These settings range from the primitive activi-

¹⁶⁶ This fact and its implications have been explored in significant depth in MICHAEL E. BRATMAN, *INTENTIONS, PLANS, AND PRACTICAL REASON* (1987). As Professor Bratman observes:

We do not [act on the basis of] plans that specify, once and for all, everything we are to do in the future. Such total plans are obviously beyond our limits. Rather, we typically settle on plans that are *partial* and then fill them in as need be and as time goes by. This characteris-

ties of hunting, battle, and exploration to more recent additions to the human behavioral repertoire such as nautical navigation, mergers and acquisitions, recreational mountain climbing, mining operations, vacation traveling, architectural design and construction projects, and even the calling of plays in football. In all these settings we employ virtually identical criteria in responding to the overlooked and unanticipated as they present themselves. The criteria are, in large part, those of means rationality—criteria for effectively accomplishing purposes in accordance with their relative priorities—with occasional reference to ends rationality in connection with the scopes of ends. The result is that there is associated with these settings a virtually universal and uniform set of intentions about addressing overlooked and unanticipated situations as they are encountered. Further, because we all address such situations in virtually the same ways with virtually identical intentions, it is often obvious how to do so and, in any event, generally feasible for us to do so on behalf of others in ways that either in fact comport with their purposes or can reasonably be regarded as comporting with their purposes. That is why agents and subordinates can successfully act for their principals and implement directives and carry out orders in the face of overlooked and unanticipated situations when the principals are not available for further direction.

It would appear that legislators do not fully appreciate the practical problems they face in formulating statutory rule systems or, presumably, they would explain in legislative history, the text of the statute being formulated, or in a special statute governing the interpretation of statutes generally, the intentions they have regarding the manner in which overlooked and unanticipated situations are to be addressed when

tic of *incompleteness* of our plans is of the first importance. It creates the need for a kind of reasoning [practical reasoning] characteristic of planning agents: reasoning that takes initial, partial plans as given and aims at filling them in with specifications of appropriate means, preliminary steps, or just relatively more specific courses of action.

Id. at 3 (emphasis in the original). See also MICHAEL E. BRATMAN, *FACES OF INTENTION: SELECTED ESSAYS ON INTENTION AND AGENCY* (1999).

they ultimately surface. Nonetheless, there are ascertainable intentions with quite specific content that can be, rationally must be, attributed to the legislature and that dictate how to address such situations. Since these intentions are implicit in human behavior, although not expressly declared by the legislature, they may in the legislative setting be called "implicit legislative intentions." They enable courts to carry out the tasks that constitute the interpretation of statutes in ways that either are or may most reasonably be viewed as intended by the legislature. They enable courts to address overlooked and unanticipated situations in consonance with the constitutional provision for legislative supremacy even when the results are not *ex ante* provided for by statutory text.

The set of implicit legislative intentions is both sophisticated and moderately numerous. It encompasses, for example, the intentions to include within a statute's coverage overlooked and unanticipated non-conceptual and substantial equivalents of paradigmatic statutory facts, to add to statutory rule systems missing substantive and remedial rules that are necessary or beneficial in giving effect to a statute's purposes, to deny effect to the presumptive semantic content of statutes when giving the presumptive semantic content effect would yield results counter to the purposes underlying the statute, to deny effect to overlooked and unanticipated successive applications of rules that produce results at variance with the underlying purposes of the statute or other higher priority purposes the law is simultaneously pursuing, to give effect to higher priority purposes when purposes of the law conflict, and to modify the meanings of words and phrases to bring their meanings into line with the purposes underlying the statute in which they appear and other purposes simultaneously operative in the law. Each of these examples reflects principles of rationality in the completion of purposive plans and designs as overlooked and unanticipated matters arise. They are, therefore, in large part objective and politically and ideologically neutral. They apply to both liberal and conservative as well as consensus legislation. They operate not to slant results politically or ideologically but appropriately to complete statutory

rule schemes and appropriately to modify meanings of words and phrases used in statutory rule schemes in light of the purposes underlying a statute and other purposes the law is simultaneously pursuing.

Not surprisingly, courts regularly and instinctively resort to these implicit legislative intentions in the interpretive process. They account for, and justify, numerous judicial decisions, many of which textualists condemn. They also account for Justice Scalia's observation that on some occasions in interpreting statutes we resort to what appear to be "objectified intentions."¹⁶⁷ These intentions are objectified in the sense that they are so inter-subjectively universal that we can virtually always count on them prevailing. It takes special evidence to convince us that they are not operative on a given occasion.

Because jurists have focused on the obvious but superficial problem of resolving an ambiguous ambiguity and not on the fundamental problems of clarifying and refining statutory purposes, completing statutory rule schemes in light of their underlying purposes and other purposes simultaneously operating in the law, and modifying the meanings of words and phrases to bring them into line with those purposes as the overlooked and unanticipated present themselves, jurists have, in the traditional theories of legislative intent and statutory purpose, overlooked the set of intentions we virtually universally employ in dealing with problems of the types we encounter in formulating statutes. Some scholars have correctly pointed out the deficiencies in the traditional notions of legislative intent and statutory purpose when we encounter situations that the legislature overlooked or failed to anticipate.¹⁶⁸ In such situations, there are, of course, no express legislative intentions. The set of implicit legislative intentions remedies those deficiencies by picking up at exactly the point at which traditional approaches to legislative intent and statutory purpose cease to be effective. By looking beyond legisla-

¹⁶⁷ SCALIA, *supra* note 4, at 17.

¹⁶⁸ See, e.g., William N. Eskridge, Jr. & Philip P. Frickey, *Statutory Interpretation as Practical Reasoning*, 42 STAN. L. REV. 321 (1990).

tive history and statutory text and taking the set of implicit legislative intentions into account, we can construct a far more realistic and complete picture of legislative intentions and purposes than jurists have traditionally recognized and employed in statutory interpretation. We can thereby address those situations that the traditional views of legislative intent and statutory purpose cannot reach.

The insights in Parts III and IV of this article alone should come close to putting textualism permanently to rest. When, in the article that is the companion to this one, we add to the mix insights from epistemology and the linguistic discipline of pragmatics, there is no doubt that textualism can be philosophically, scientifically, and linguistically eliminated as an intellectual force in the law.¹⁶⁹ Further, the set of implicit legislative intentions, together with the linguistic discipline of pragmatics, will also allow us permanently to set aside many of the canons of construction and appropriately to revise and justify those that have some continuing validity. They will also provide a much needed answer to the excessive claims of nihilistic skeptics. As an added bonus, when we employ them, the results we achieve are epistemologically unassailable. And, somewhat surprisingly, the linguistic discipline of pragmatics makes clear that employing them in the interpretive process accords with the cognitive processes whereby we comprehend language generally.

When the implicit legislative intentions are identified in full and applied in statutory interpretation, the problems we encounter in interpreting statutes shrink to quite manageable proportions, although they do not vanish entirely. We cannot achieve the ideal. Nonetheless, this set of intentions allows for a much higher percentage of determinate, albeit, not always certain, outcomes than does the usable but theoretically inad-

¹⁶⁹ Textualism operates on the assumption that language is simply a code. Note 17, *supra*, indicates why that is wrong. The discussion above also makes it clear why textualism creates such a drag on remedial legislation. It prevents courts from acting in overlooked and unanticipated situations to rationally complete such legislation in accordance with its purposes.

equate and partially conflicting mishmash of traditional legal reasoning.¹⁷⁰ We see that many decisions in the reporters were inevitable in light of these intentions. There are, however, residual situations, generally at the periphery of statutory rule systems, in which even the set of implicit legislative intentions does not mandate entirely objective, politically and ideologically neutral, determinate outcomes. There are situations in which the set of implicit legislative intentions can do no more than mandate that judges, parties who ideally have no personal stake in the outcome, attempt to provide a rational, fair, and just outcome as best they can conceive one. That, however, is an imperfection that we can live with. Some might rationally conclude that it is not even an imperfection.

¹⁷⁰ Outcomes that, on the available evidence, are favored by the probabilities, for example, that the legislature intended them, cannot be rejected and are, therefore, determinate. Whether the legislature actually intended them can, nonetheless, be uncertain. The same holds in gambling. Once one knows the odds, the proper bet is fixed. The outcome on a specific occasion may, however, render a proper bet a losing bet. Nonetheless, we rationally act in all settings on the basis of the probabilities.

AMERICA'S ANTI-STANDING ARMY TRADITION AND THE SEPARATE COMMUNITY DOCTRINE

*Earl F. Martin**

INTRODUCTION

A critical element in America's struggle to fashion a civil-military relationship that serves both its democratic principles and its security needs is the imperative of striking a proper balance between the rights of individual service members and the needs of the country's military establishment. On one hand, we have individuals possessed of constitutionally protected rights, while on the other, we have an institution tasked to defend the country's national security. As might be expected, this has inevitably led to a clash between individual rights versus military efficiency, and this clash has led to litigation. The U.S. Supreme Court's response to this litigation has been to deploy the "separate community doctrine."

The separate community doctrine is the standard of review that the Supreme Court employs when called upon to settle disputes between military members and the armed forces.¹ The doctrine is most likely to be invoked when a conflict revolves around a claim that the military has in one way or another violated a service member's constitutional rights. These cases generally pit a claim that these rights are being violated against the claim that the armed forces need to proceed in a particular way in order to accomplish their mission.

* Dean and Professor of Law, Gonzaga University School of Law. B.A. & J.D., University of Kentucky; LL.M., Yale Law School. The author would like to thank Justin T. Cunningham for his valuable research assistance.

¹ Earl F. Martin, *Separating United States Service Members from the Bill of Rights*, 54 SYRACUSE L. REV. 599, 601 (2004).

Although language anticipating the separate community doctrine can be found in Supreme Court cases from the 1800s,² it was not until the 1950s that the doctrine emerged as a recurring influence in the Court's jurisprudence. Three cases from that decade, *Orloff v. Willoughby*,³ *Burns v. Wilson*,⁴ and *United States ex rel. Toth v. Quarles*,⁵ served to provide a foundation upon which the Court would fashion its approach to deciding disputes between the armed forces and individual soldiers, sailors, and airmen. Each of these cases contains quotes that have been recited time and again by the Court as it has embraced the proposition that the civilian courts should tread lightly when asked to interfere with action taken by the military against one of its own.⁶ Over the years, the Court has borrowed language from each to proclaim that as "a specialized community governed by a separate discipline from that of the civilian,"⁷ the military is subject to "certain overriding demands of discipline and duty"⁸ and has the responsibility "to fight or be ready to fight wars should the occasion arise."⁹

The separate community doctrine has evolved since its birth in the 1950s. After the foundation cases of that decade, the doctrine lay relatively dormant at the Supreme Court

² See, e.g., *Martin v. Mott*, 25 U.S. 19, 30 (1827) (speaking of the need for service members to respond with "efficient and immediate compliance" to military commands so as not to "jeopard[ize] the public interests"); *United States v. Grimley*, 137 U.S. 147, 153 (1890) ("An army is not a deliberative body. It is the executive arm. Its law is that of obedience. No question can be left open as to the right to command in the officer, or the duty of obedience in the soldier.").

³ 345 U.S. 83 (1953).

⁴ 346 U.S. 137 (1953).

⁵ 350 U.S. 11 (1955).

⁶ Notwithstanding the Court's liberal use of the foundation cases in this effort, it is debatable whether these cases should be carrying so much weight in support of a strong application of the separate community doctrine. See *Martin*, *supra* note 1, at 615-18.

⁷ See, e.g., *Parker v. Levy*, 417 U.S. 733, 744 (1974) (quoting *Orloff v. Willoughby*, 345 U.S. 83, 94 (1953)).

⁸ See, e.g., *Middendorf v. Henry*, 425 U.S. 25, 43 (1976) (quoting *Burns v. Wilson*, 346 U.S. 137, 140 (1953)).

⁹ See, e.g., *Rostker v. Goldberg*, 453 U.S. 57, 70 (1981) (quoting *U.S. ex rel. Toth v. Quarles*, 350 U.S. 11, 17 (1955)).

until the 1969 case of *O'Callahan v. Parker*.¹⁰ Employing a weak form of the doctrine as characterized by a reticence on the part of the Court to grant deference to the military establishment, the *O'Callahan* Court used the uniqueness of the military community as justification to restrict the reach of court-martial jurisdiction in order to protect service members' Fifth and Sixth Amendment jury rights, as those rights do not have force in military courts.¹¹ Following *O'Callahan*, the Court entered a period in which a moderate form of the separate community doctrine was ascendent.¹² From the 1974 case of *Parker v. Levy*¹³ to the 1983 case of *Chappell v. Wallace*,¹⁴ the Court used the separate community doctrine to inform, but not dominate, its decisions.¹⁵ For example, in

¹⁰ 395 U.S. 258 (1969).

¹¹ Martin, *supra* note 1, at 618-22. The Fifth and Sixth Amendment jury rights do not have force in military courts. *O'Callahan*, 395 U.S. at 273. The Fifth Amendment to the U.S. Constitution provides: "No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury, *except in cases arising in the land or naval forces*, or in the Militia, when in actual service in time of War or public danger" U.S. CONST. amend. V. (emphasis added). Although the excepting language in this provision could be read to mean that "cases arising in the land or naval forces," as well as in the militia, are only excepted from the requirement of grand jury presentment or indictment "in actual service in time of War or public danger," the Court has interpreted the "time of War" provision as referring only to cases arising in the militia, not the land or naval forces. *Id.*; see, e.g., *Johnson v. Sayre*, 158 U.S. 109, 114 (1895); *Ex parte Mason*, 105 U.S. 696, 701 (1881); *Dynes v. Hoover*, 61 U.S. 65, 78 (1857). The Supreme Court held in 1866 that the Sixth Amendment right to a jury trial does not apply in court-martial proceedings. *Ex parte Milligan*, 71 U.S. (4 Wall.) 2, 123 (1866). The *Milligan* Court stated that "the framers of the Constitution, doubtless, meant to limit the right of trial by jury, in the sixth amendment, to those persons who were subject to indictment or presentment in the fifth." *Id.*; see also *Ex parte Quirin*, 317 U.S. 1, 40 (1942) ("[C]ases arising in the land or naval forces' . . . are expressly excepted from the Fifth Amendment, and are deemed excepted by implication from the Sixth.") (citing *Ex parte Milligan*, 71 U.S. (4 Wall.) at 123, 138, 139).

¹² See Martin, *supra* note 1, at 622-42.

¹³ 417 U.S. 733 (1974).

¹⁴ 462 U.S. 296 (1983).

¹⁵ A complete list of the moderate separate community cases includes: *Chappell v. Wallace*, 462 U.S. 296 (1983); *Rostker v. Goldberg*, 453 U.S. 57 (1981); *Sec'y of Navy v. Huff*, 444 U.S. 453 (1980); *Brown v. Glines*, 444 U.S. 348 (1980); *Greer v. Spock*, 424 U.S. 828 (1976); *Schlesinger v. Councilman*, 420 U.S.

Chappell, the separate community doctrine served as a "special factor" in the context of a pre-existing test which had anticipated that the unique facts of any given case might counsel in favor of restricting the military establishment's liability for harm done to individual service members.¹⁶

Since 1983, a strong form of the separate community doctrine has become the Supreme Court's standard of choice. This form, which grants an excessive amount of deference to the actions and decisions of the military establishment in its disputes with individual service members, made its first appearance in the 1976 case of *Middendorf v. Henry*.¹⁷ It has been the sole form of the doctrine employed by the Court since the 1986 case of *Goldman v. Weinberger*.¹⁸ The evolution of the separate community doctrine into its strong form involved the fusion of two ideas under this common title. The first is the notion that is captured in the quotes recreated above from the 1950s foundational cases.¹⁹ This is the idea that the uniqueness of the military community justifies subjecting service members to different standards than similarly situated civilians when it comes to the application of the guarantees in the Constitution and the Bill of Rights.²⁰ The second idea that props up the strong application of the separate community doctrine is the claim that the Constitution gives Congress almost unfettered power over the regulation and management of the armed forces.²¹ This assertion has been a part of most

738 (1975); *Schlesinger v. Ballard*, 419 U.S. 498 (1975); *Parker v. Levy*, 417 U.S. 733 (1974). See Martin, *supra* note 1, at 622-42.

¹⁶ 462 U.S. at 300-04; see also Martin, *supra* note 1, at 640-42.

¹⁷ 425 U.S. 25 (1976).

¹⁸ 475 U.S. 503 (1986).

¹⁹ See *supra* notes 7-9 and accompanying text.

²⁰ See, e.g., *Goldman*, 475 U.S. at 506-07 (refusing to apply the then-prevailing standard from *Sherbert v. Verner*, 374 U.S. 398, 406 (1963), for free exercise of religion claims to a case involving an Orthodox Jew and ordained rabbi who was ordered not to wear his yarmulke while he was serving in uniform as an Air Force officer). One commentator has referred to the *Goldman* majority opinion as representing "a full flowering of the notion . . . that the military is a separate community subject to separate norms different from those of civilian life." Stephanie A. Levin, *The Deference that Is Due: Rethinking the Jurisprudence of Judicial Deference to the Military*, 35 VILL. L. REV. 1009, 1020 (1990).

²¹ See, e.g., *United States v. Stanley*, 483 U.S. 669, 681-84 (1987) (viewing

of the separate community doctrine cases over the years, but it has become ever more influential as the Court has continued to refine the doctrine.²²

Embracing the adage that you never know where you are going until you know where you have been, this Article explores the relationship that existed at the time of the country's founding between the military establishment, individual service members, and the Constitution as part of an effort to test the legitimacy of the Supreme Court's separate community doctrine jurisprudence. The lens through which this exercise will be undertaken is the debate that raged at the founding over the national government raising and maintaining a standing army.²³ Between the years leading up to the War for Independence and the end of the eighteenth century, the perceived benefits and dangers presented by a national standing army dominated the discussion of how to meet the country's security needs.²⁴ Therefore, what follows is a consideration of what was being said in America about standing armies in general, and a national standing army in particular, from the 1760s to the end of the 1790s. The material reviewed

Congress's plenary power over the armed forces as a "special factor[] counseling hesitation" against allowing a former enlisted man to recover damages from superior officers for having secretly administered lysergic acid diethylamide (LSD) to him as part of an Army effort to study the effects of the drug on human subjects) (citation omitted).

²² See, e.g., *Weiss v. United States*, 510 U.S. 163, 177 (1994) (explaining that although the Due Process Clause provides some measure of protection to defendants in military proceedings, "in determining what process is due, courts 'must give particular deference to . . . determination[s] of Congress, made under its authority to regulate the land and naval forces'"); *Solorio v. United States*, 483 U.S. 435, 450-51 (1987) (holding that Congress's plenary power over the armed forces gives it the authority to make the military status of the offender the determining factor for delineating the scope of court-martial jurisdiction).

²³ "The phrase 'standing army' seems to have arisen from the contrast between the images of a 'marching' and a 'standing' army: that is, between one 'taking the field against a real enemy or present danger' and one not actively engaged in necessary military activity, but literally 'standing around.'" Levin, *supra* note 20, at 1024 (citing J.G.A. Pocock, *THE MACHIAVELLIAN MOMENT: FLORENTINE POLITICAL THOUGHT AND THE ATLANTIC REPUBLICAN TRADITION* 411 (1975)).

²⁴ See RICHARD H. KOHN, *EAGLE AND SWORD* 1-13 (1975).

comes from the Revolutionary War era (Part I), from the time of the framing and adoption of the Constitution and the Bill of Rights (Part II), and from the last decade of the eighteenth century as the country struggled to create and manage a federal military establishment consistent with the demands of the new Constitution (Part III). This article will conclude with a discussion of the first principles for the management of the country's military establishment that came out of the founding generation's struggle with the controversy of a national standing army (Part IV).

Before proceeding with the just promised review of history, two points of clarification and one caveat need to be offered. The first point of clarification is to note that the sought after historical knowledge is not intended to be an end unto itself. To the extent we are able to arrive at some appreciation of what the founding generation thought about the propriety of standing armies, that knowledge will then be used to inform our understanding of the proper relationship that should exist today between the military establishment, individual service members, and the Constitution.²⁵

The second point of clarification is that the phrase "founding generation" has been deliberately chosen to describe whose point of view is being sought in this work. Starting from the specific and progressing to the general, our attention could be directed to what a particular person from the time of the country's founding thought about the propriety of standing armies, or what the Framers of the Constitution thought about the same, or the ratifiers of that document, or the population of the country as a whole from the late eighteenth century.²⁶ Indeed, as we consider the issue of a national

²⁵ See generally H. Jefferson Powell, *Rules for Originalists*, 73 VA. L. REV. 659, 666-67 (1987) ("[H]istorical knowledge must not become an end in itself rather than a means. . . . The turn to history is legitimate only so long as it subserves the interpretation of, and fidelity to, the Constitution.").

²⁶ The term "Framers" is meant to represent those persons "who helped to draft the text that the Constitutional Convention proposed to the states for ratification." 1 LAURENCE H. TRIBE, *AMERICAN CONSTITUTIONAL LAW* § 1-14, at 57 (3d ed. 2000). The term "ratifiers" is meant to represent those legislators who actual-

standing army, we will, at various points, consider comments on the matter made by various individuals or groups who could be said to represent all of the above. However, the point of view coveted here is that of those members of the country from the time of its founding who were writing and speaking about standing armies, and those whom these writers and speakers expected (or hoped) to reach. In this sense, we are after the collective opinion of the politically active and/or attentive members of late 1700s America; *i.e.*, the polity.

The Federalist Papers serve as a nice example of a source of information that will be mined in search of the collective opinion that is being sought in this article. The letters of Alexander Hamilton, James Madison, and John Jay, published in New York newspapers between 27 October 1787 and 28 May 1788,²⁷ are an excellent source of information, not only for what the authors of those letters thought about the critical issues surrounding the ratification of the newly drafted Constitution, but also for the broader public opinion regarding those same concerns.²⁸ Although it has been argued that *The Federalist Papers* were "propaganda pieces that . . . were tintured by the 'zeal of advocates,'"²⁹ and thus, should be read

ly ratified the Constitution in the various state conventions. Paul Brest, *The Misconceived Quest for the Original Understanding*, 60 B.U. L. REV. 204, 214 (1980).

²⁷ Richard B. Morris, *The Origin and Framing of the American Constitution*, in *THE UNITED STATES MILITARY UNDER THE CONSTITUTION OF THE UNITED STATES, 1789-1989*, 41, at 55 (Richard H. Kohn ed., 1991).

²⁸ See generally ANTONIN SCALIA, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, in *A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW* 38 (Amy Gutmann ed., 1997) ("I will consult the writings of some men who happened to be delegates to the Constitutional Convention—Hamilton's and Madison's writings in *The Federalist*, for example. I do so, however, not because they were Framers and therefore their intent is authoritative and must be the law; but rather because their writings, like those of other intelligent and informed people of the time, display how the text of the Constitution was originally understood.").

²⁹ Powell, *supra* note 25, at 681 (quoting Letter from James Madison to Edward Livingston (Apr. 17, 1824), *reprinted in* 3 *LETTERS AND OTHER WRITINGS OF JAMES MADISON* 435, 436 (Philadelphia 1865)); see also GEOFFREY R. STONE ET AL., *CONSTITUTIONAL LAW* 14 (5th ed. 2005) ("Although [*The Federalist Papers*] are often consulted as a means of understanding the theory underlying the Constitution, the 'intentions' of its drafters, it is important to keep in mind that the es-

with some amount of circumspection, this does not mean that their content lacks legitimacy in helping us understand where the polity stood on various issues of the day.³⁰ Even if Alexander Hamilton, for example, was advocating positions in his letters that he did not personally hold,³¹ the very fact that he advanced those positions in order to win support for the Constitution tells us where he believed the public stood on the issues surrounding ratification.³² In this way, the content of *The Federalist Papers* is some reflection of the opinions held by the politically active and attentive members of late eighteenth century America.³³

says were in many respects propaganda pieces, designed to persuade the ambivalent.”).

³⁰ See Seth Barrett Tillman, *The Federalist Papers as Reliable Historical Source Material for Constitutional Interpretation*, 105 W. VA. L. REV. 601, 617 (2003).

The minds of the three authors of *The Federalist Papers* were occupied by the great issue of their day, the issue that would occupy our nation's citizens and statesmen through the Revolution, up to and through the irrepressible conflict and into Reconstruction, and perhaps beyond. That issue was Union. Beyond that *The Federalist Papers* fails as a clause by clause detailed defense or explication of the Constitution of 1787. And that is as it should be. The purpose of its publication was not to furnish a resource in the legal wrangling and law review nitpicking of our generation, but rather to encourage ratification in theirs.

Id. (footnote omitted).

³¹ See, e.g., Powell, *supra* note 25, at 681 (arguing that Hamilton's overall ideas about federalism and his subsequent behavior call into question the sincerity of his claim in *The Federalist Number 81* that Article III to the Constitution would not strip away sovereign immunity created by state law).

³² See generally *id.* at 687 (noting that it has often been suggested that the Constitution's supporters “frequently exaggerated the degree of autonomy they expected and intended the Constitution to leave to the states in a successful effort to allay localist fears”).

³³ See generally ROBERT H. BORK, *THE TEMPTING OF AMERICA: THE POLITICAL SEDUCTION OF THE LAW* (1990). The point being argued is analogous to the point Judge Bork made regarding the search for subjective intentions in constitutional interpretation, to wit:

[W]hat the ratifiers understood themselves to be enacting must be taken to be what the public of that time would have understood the words to mean. It is important to be clear about this. The search is not for a subjective intention. If someone found a letter from George Washington to Martha telling her that what he meant by the power to lay taxes was

At this point it could be said that the task described in the last two paragraphs is not capable of being accomplished given that the founding generation as defined above is too large and too diverse to be said to have held a singular position on the question of standing armies.³⁴ Because this criticism is true enough, it becomes necessary to deliver the previously promised caveat. The goal herein is not to say, for example, that thirteen percent of the founding generation supported the proposition of having a standing army, while eighty-four percent were opposed to the same and three percent undecided. That type of precision, or anything approaching it, is not possible to achieve at this point in time, if ever it was. Put another way, the goal is not to be able to say that history proves that the founding generation either favored or disfavored standing armies, as historical inquiry is rarely able to "prove" such claims to the point of rendering intelligent disagreement impossible.³⁵ Instead, the objective is to arrive at an understanding of the broad social consensus, if any, that existed within the polity of early America regarding the place of the military establishment within our constitutional scheme of government.³⁶

not what other people meant, that would not change our reading of the Constitution in the slightest. Nor would the subjective intentions of all the members of a ratifying convention alter anything. When lawmakers use words, the law that results is what those words ordinarily mean.

Id. at 144.

³⁴ See Ronald Dworkin, *The Forum of Principle*, 56 N.Y.U. L. REV. 469, 477 (1981) ("[T]here are no, or very few, relevant collective intentions, or perhaps only collective intentions that are indeterminate rather than decisive one way or another."). See generally Powell, *supra* note 25, at 663 (referring to "the thorny issue of group intent").

³⁵ Powell, *supra* note 25, at 678.

³⁶ See generally Brest, *supra* note 26, at 214.

Interpreters often treat the writings or statements of the framers of a provision as evidence of the adopters' intent. This is a justifiable strategy for the moderate originalist who is concerned with the framers' intent on a relatively abstract level of generality—abstract enough to permit the inference that it reflects a broad social consensus rather than notions peculiar to a handful of the adopters. It is a problematic strategy for the strict originalist.

I. THE REVOLUTIONARY WAR ERA

American colonists of the 1760s and 1770s borrowed heavily from seventeenth and eighteenth century radical English Whig arguments in scripting their challenge to the British standing military establishment in North America.³⁷ Even though opposition to permanent military forces was politically unrealistic in England at least as early as 1690 due to external threats to that country's security, that position remained a constitutional imperative in the home country well into the 1700s.³⁸ Protests against Britain's military establishment were frequently heard in Parliament, often in association with the annual renewal of the Mutiny Act, thus forcing the government to continually defend the constitutionality of those forces.³⁹

English opposition to a standing military generally coalesced around the objection that any force large enough to protect the country from external threats would present an untenable danger to individual liberties.⁴⁰ Viewing England's

Id.

³⁷ See JOYCE LEE MALCOLM, *TO KEEP AND BEAR ARMS* 142-47 (1994); Jerry M. Cooper, *Federal Military Intervention in Domestic Disorders*, in *THE UNITED STATES MILITARY UNDER THE CONSTITUTION OF THE UNITED STATES, 1789-1989*, 120, at 132 (Richard H. Kohn ed., 1991); JOHN PHILLIP REID, *IN DEFIANCE OF THE LAW* 4, 31 (1981); Levin, *supra* note 20, at 1026-29. Professor Fisher has described the "Radical Whigs" of British North America, circa 1776, as follows:

[They] were less fearful of commerce than of standing armies, patronage, and the spread of 'luxury.' In contrast to their Tory cousins, they believed that the economic independence essential to responsible citizenship could and should be achieved by a large percentage of the male population. They were also more likely to subscribe to the collection of beliefs commonly known as civic humanism — the notions that a good life is a virtuous life; that virtue consists partly in a willingness to subordinate one's private ends to the commonwealth; and that 'man [is] a political being whose realization of self occurs only through participation in public life, through active citizenship in a republic.'

William W. Fisher III, *Ideology, Religion, and the Constitutional Protection of Private Property: 1760-1860*, 39 *EMORY L.J.* 65, 74 (1990) (alteration in original) (footnote omitted).

³⁸ REID, *supra* note 37, at 85-89, 92-93.

³⁹ *Id.* at 92-93.

⁴⁰ *Id.* at 93-94.

constitution as a product of consensus and popular custom versus sovereignty and power, John Trenchard, a political commentator and member of the Whig reform party, warned in 1697 that if the government continued to execute law with military force "the Constitution must either break the Army, or the Army will destroy the Constitution."⁴¹ Twenty-four years later, in one of his *Cato* letters, Trenchard elaborated upon this sentiment when he claimed "that all parts of Europe which are enslaved, have been enslaved by armies; and it is absolutely impossible, that any nation which keeps them amongst themselves can long preserve their liberties; nor can any nation perfectly lose their liberties who are without such guests."⁴² Trenchard's arguments were carried forward in 1739, when Lord Bathurst declared in the House of Lords that a standing army large enough to protect the country would "be sufficient for subduing and enslaving the people, as soon as their superiors shall give them the word of command."⁴³

In the face of the dangers presented by a standing army, many in England offered the militia, the home guard generally made up of all the adult male citizens, as an alternative

⁴¹ See *id.* at 108. Also writing in 1697, the Reverend Samuel Johnson declared that no king could be trusted with a standing army even "if he were an Angel." See *id.* at 87 (quoting SAMUEL JOHNSON, *THE WORKS OF THE LATE REVEREND MR. SAMUEL JOHNSON: SOMETIME CHAPLAIN TO THE RIGHT HONOURABLE WILLIAM, LORD RUSSELL* 325 (2d ed. 1713)).

⁴² JOHN TRENCHARD & THOMAS GORDON, 2 *CATO'S LETTERS* 682 (Ronald Hamowy ed., 1995); see also REID, *supra* note 37, at 156 (Members of the House of Lords protested a vote in 1718 to maintain a peacetime army of 16,347 men by saying that "so numerous [a] force . . . may, we fear, endanger our constitution, which hath never yet been entirely subverted but by a standing army.") (quoting 1 *A COMPLETE COLLECTION OF THE PROTESTS OF THE LORDS WITH HISTORICAL INTRODUCTIONS* 241 (James E. Thorold Rogers ed., 1875)).

⁴³ REID, *supra* note 37, at 94 (quoting Speech of Lord Bathurst, *Lords Debates of 9 March 1739*, in 10 *PARLIAMENTARY HISTORY OF ENGLAND FROM THE EARLIEST PERIOD TO THE YEAR 1803*, at 543 (1806-1820)); see also *id.* at 6 (Philip Yorke, the Earl of Hardwicke, in a speech before the House of Lords in May of 1756, endorsed the practice of attaching a preamble to the annual Mutiny Bill which stated, "That a standing army in time of peace; execution of martial law in time of peace, are against the law.") (quoting Speech of Lord Chancellor Hardwicke, *Lords Debates of 24 May 1756*, in 15 *PARLIAMENTARY HISTORY* at 727-28).

national security force.⁴⁴ The militia's proponents argued that it "had defended England in better days and could do so again."⁴⁵ It was not that these supporters actually believed that the militia could defend the country against foreign invasion as competently as professional soldiers, as most conceded that the latter had the advantage in warfare;⁴⁶ rather, the pro-militia crowd believed that the military advantage offered by standing armies was outweighed by the danger those forces presented to civil freedoms.⁴⁷ The Earl of Shaftesbury, Anthony Ashley Cooper, captured the sentiment perfectly in 1675 when he wrote, "The *Militia* must, and can never be otherwise than for *English Liberty*, 'cause else it doth destroy it *self*, but a *standing Force* can be nothing but *Prerogative*, by whom it hath its *idle Living* and *Subsistence*."⁴⁸ Fifty-eight years later, the standing army antagonist Lord Bathurst echoed the Earl of Shaftesbury when he argued in Parliament that "[t]he militia of the kingdom are the people of the kingdom and it is impossible to make use of the people for oppressing the liberties of the people."⁴⁹

Although Parliamentary supremacy and its maintenance of a standing military force had become part of the British constitution by the time revolution was fermenting in America, declarations against standing armies remained a part of the English political discourse.⁵⁰ A correspondent in *The London Magazine* in 1769 sided with the citizens of Boston against the ministers who had ordered the British army to that city by writing that "[t]he same oppressors under which we groan, are their oppressors . . . the same freedom which is

⁴⁴ See *id.* at 101-04.

⁴⁵ *Id.* at 103.

⁴⁶ *Id.*

⁴⁷ *Id.*

⁴⁸ *Id.* at 102 (quoting Letter from Anthony Ashley Cooper, Earl of Shaftesbury. A Letter from a Parliamentman to his Friend, Concerning the Proceedings of the House of Commons this last Sessions (Oct. 13, 1675)).

⁴⁹ *Id.* at 104 (quoting Speech of Lord Bathurst, *Lords Debates of 6 March 1733*, in 8 PARLIAMENTARY HISTORY OF ENGLAND FROM THE EARLIEST PERIOD TO THE YEAR 1803, at 1249 (1806-1820)).

⁵⁰ See *id.* at 98-99.

the object of our generous pursuit, is the sole object of theirs.”⁵¹ George Johnstone, who had served as the Governor of West Florida, spoke more directly to the danger of a standing army in a debate in the House of Commons in 1774 when he said that “the liberties of [the] country must, in the end, fall a sacrifice to that power which has annihilated the rights of mankind in other states.”⁵² One year later, just as the Revolutionary War was breaking out in America, the English Whig pamphleteer James Burgh wrote in his influential work *Political Disquisitions* that “[i]n a survey of public abuses, it would be unpardonable to overlook that of a standing army in times of peace, one of the most hurtful, and most dangerous of abuses.”⁵³

Few of the rights traditionally associated with the ancient English constitution of consensus and popular custom were better known to Americans of the 1760s and 1770s than the right to be free of standing armies.⁵⁴ Accordingly, the rhetoric of American patriots in the years leading up to the War for Independence borrowed heavily from their predecessors and contemporaries in England by highlighting the threat of a standing army being used to impose and support dictators and deny civil rights.⁵⁵ In the summer before the Stamp Act crisis erupted at the end of 1765, the *Providence Gazette* warned that “[b]road hints have been given that *standing forces* are to be sent amongst us, to humble us, and to enforce execution of such laws, as we must esteem grievances.”⁵⁶ A few years lat-

⁵¹ See *id.* at 31-32 (quoting 38 *The London Magazine* 205 (1769)). The correspondent was no doubt motivated by bloodshed that had resulted from the use of the British military to enforce civil law and suppress rioters in England. See *id.* at 117-19.

⁵² See *id.* at 10, n.14 (quoting Speech of Johnstone, *Commons Debates of 16 December 1774*, in 4 *AMERICAN ARCHIVES* 1480 (Peter Force ed., 1837-1853)).

⁵³ 2 JAMES BURGH, *POLITICAL DISQUISITIONS: OR, AN ENQUIRY INTO PUBLIC ERRORS, DEFECTS, AND ABUSES* 341 (1775).

⁵⁴ REID, *supra* note 37, at 80.

⁵⁵ See *supra* note 37 and accompanying text. English Whig pamphleteers like John Trenchard and James Burgh were prime sources of influence in this regard. See KOHN, *EAGLE*, *supra* note 24, at 4.

⁵⁶ See REID, *supra* note 37, at 72 (quoting BOSTON POST-BOY, July 15, 1765, at 2). The Stamp Act, which was passed by Parliament in March of 1765, “levied

er, following the landing of British troops in Boston on October 1, 1768,⁵⁷ Samuel Adams attacked the British government's use of its military forces in a series of letters published in the *Boston Gazette*.⁵⁸ In one letter dated December 12, 1768, Adams claimed that "it [was] a very improbable supposition, that any people can long remain free, with a strong military power in the very heart of their country,"⁵⁹ while in another published in February of 1769, he wrote about "the exercise of . . . military power [being] forever *dangerous* to civil rights."⁶⁰ In 1774, John Hancock, in his oration on the Boston Massacre, channeled the Earl of Shaftsbury from a hundred years earlier when he contrasted a well-regulated militia with the British regulars who were policing the colonies.⁶¹ While the latter had "trampl[ed] on the rights and

a tax . . . on legal documents, almanacs, newspapers, and nearly every form of paper used in the colonies." 1 BERNARD BAILYN ET. AL., *THE GREAT REPUBLIC: A HISTORY OF THE AMERICAN PEOPLE* 229 (4th ed. 1992). The following October, "thirty-seven delegates from nine colonies met in New York in the Stamp Act Congress and drew up a set of formal declarations and petitions denying Parliament's right to tax them." *Id.* at 231. The crisis over the Stamp Act "stimulated bold political and constitutional writings throughout the colonies, deepened [the colonists'] political participation and consciousness, and produced new forms of organized [popular] resistance." *Id.* at 233; see also DONALD M. SNOW & DENNIS M. DREW, *FROM LEXINGTON TO DESERT STORM AND BEYOND* 35 (2d ed. 2000) (noting that the Quartering Act of June 1765, which required citizens to open their homes to British troops, led to a cry against quartering being heard across New England).

⁵⁷ See BAILYN ET. AL., *supra* note 56, at 236. By 1769 there were nearly 4000 armed redcoats in Boston, which at the time had a total population of only 15,000. *Id.*

⁵⁸ See I THE WRITINGS OF SAMUEL ADAMS 134-447 (Harry Alonzo Cushing ed., 1904) (collecting Adams's letters from 1768-1769).

⁵⁹ *Id.* at 264. This letter was signed "Vincex." *Id.* In a December 26 letter, Adams wrote of a corrupt minister using the military to "subdue the people of America to his schemes of arbitrary and despotic power." *Id.* at 275-76.

⁶⁰ See *id.* at 318 (this letter was signed "E.A."). In a letter published in the *Public Advertiser* on February 1, 1770, Benjamin Franklin argued that "[f]ree Government depends on Opinion, not on the brutal Force of a Standing Army." REID, *supra* note 37, at 155 (quoting PUBLIC ADVERTISER, Feb 1, 1770). An unnamed correspondent to the *New Hampshire Gazette* wrote in 1770 that "[s]tanding armies have ever proved destructive to the Liberties of a People, and where they are suffered, neither Life nor Property are secure." *Id.* at 95 (quoting NEW HAMPSHIRE GAZETTE, Apr. 2, 1770, at 4).

⁶¹ John Hancock, *Boston Massacre Oration* (Mar. 5, 1774), available at

liberties of [the people],” the former was “a safe, an honorable guard to a community . . . and . . . laudably tenacious of that freedom in which they were born.”⁶² Just one month before the opening shots of the Revolutionary War were fired on Lexington Green on April 19, 1775,⁶³ Patrick Henry, in his “Give Me Liberty Or Give Me Death” speech, characterized the British fleets and armies in the colonies as “the implements of war and subjugation,” and said that the same were “the last arguments to which kings resort.”⁶⁴

The widespread American sentiment against standing armies also found expression during the revolutionary period at various levels of colonial and early American government. In a letter issued in the Town of Medford, Massachusetts, in March of 1770, warned that under British military rule the colonies would be “dragooned to an obedience to Laws made, not only without, but against their Consent.”⁶⁵ In full agree-

ahp.gatech.edu/boston_mass_orat_1774.html [hereinafter *Oration*] (delivered at Boston, Massachusetts, on the anniversary of the Boston Massacre of 1770). The Boston Massacre occurred on March 5, 1770, when British troops fired on a threatening crowd and killed five civilians. BAILYN ET. AL., *supra* note 56, at 236.

⁶² *Hancock, supra* note 61; see also Thomas Jefferson, *A Summary View of the Rights of British America* (July 1774), available at <http://press-pubs.uchicago.edu/founders/documents/v1ch14s10.html>. Jefferson asserted that if the Crown had the right to station large bodies of troops among the civilian population, then the King could “swallow up all our other rights whenever he should think proper.” *Id.*

⁶³ See BAILYN ET. AL., *supra* note 56, at 252.

⁶⁴ Patrick Henry, Speech at The Second Virginia Convention (Mar. 23, 1775), reprinted in NORINE DICKSON CAMPBELL, *PATRICK HENRY: PATRIOT AND STATESMAN* 129 (1969). Later in his speech, in an expression of his opinion of the power held by a free people, Henry said, “Three millions of people, armed in the holy cause of liberty, and in such a country as that which we possess, are invincible by any force which our enemy can send against us.” *Id.* at 130 (citation omitted).

⁶⁵ See REID, *supra* note 37, at 72 (quoting Letter from the Town of Melford to the Town of Boston, Mar. 14, 1770, in *BOSTON GAZETTE* (Supp.), Mar. 19, 1770, at 2); see also *id.* at 106 (describing the Town of Boston’s clear preference for militia over a standing army as evidenced by the town’s voting in 1773 that “the Sword should never be entrusted but to those who combat *pro aris et focis*, and whose interest is to preserve the public Peace”) (quoting *Instructions of the Town of Boston*, 5 May 1773, *GAZETTE AND POST-BOY*, May 10, 1773, at 1). In the fall of 1772, in a publication of the proceedings of their town meetings, Bostonians identified the introduction of a standing army in peacetime as one of the significant violations of American rights carried out by the British government. BAILYN

ment with its political subdivision, the Massachusetts House of Representatives proclaimed in April of 1770 that "[a] military force . . . posted among the People, without their express Consent, is . . . one of the greatest Grievances, and threatens the total subversion of a free Constitution."⁶⁶ The fear that implicitly underlay the declarations of both the Town of Medford and the Massachusetts House was explicitly stated in the 1776 Virginia Declaration of Rights.⁶⁷ In language that was repeated, in one form or another, in seven other state constitutions from the time, the Virginia Declaration affirmed that "a well regulated militia . . . is the proper, natural and safe defen[s]e of a free state [and] . . . standing armies, in time of peace, should be avoided, as dangerous to liberty."⁶⁸

Two actions taken at the national level that served as bookends to the Revolutionary War and that were inspired, in part, by the polity's distrust of standing armies were the publication of the Declaration of Independence and the adoption of the Articles of Confederation. Members of the Continental Congress specifically cited the abuses and dangers of a standing army as grounds for cessation from England when they put their signatures to the Declaration of Independence on July 4, 1776.⁶⁹ Included within the Declaration's list of "re-

ET. AL., *supra* note 56, at 239.

⁶⁶ MALCOLM, *supra* note 37, at 144. One year earlier that same body declared in a public message to Governor Francis Bernard that "[t]he experience of the ages is sufficient to convince that the military power is ever dangerous, and subversive of free constitutions." See REID, *supra* note 37, at 155-56 (quoting *Message from the Massachusetts House of Representatives to Governor Francis Bernard, 31 May 1769*, in *SPEECHES OF THE GOVERNORS OF MASSACHUSETTS FROM 1765 TO 1775; AND THE ANSWERS OF THE HOUSE OF REPRESENTATIVES TO THE SAME; WITH THEIR RESOLUTIONS AND ADDRESSES FOR THAT PERIOD 167 (1818)*).

⁶⁷ THE COMPLETE BILL OF RIGHTS: THE DRAFTS, DEBATES, SOURCES, AND ORIGINS 185 (Neil H. Cogan ed., 1997).

⁶⁸ *Id.* The other seven state constitutions that contained very similar statements were Delaware, Maryland, Massachusetts, New Hampshire, North Carolina, Pennsylvania, and Vermont. See Levin, *supra* note 20, at 1031. These state constitutions "denounced standing armies and endorsed a militia, provided that it was a general and not a select militia." MALCOLM, *supra* note 37, at 150.

⁶⁹ THE DECLARATION OF INDEPENDENCE (U.S. 1776). In drafting the Declaration, Thomas Jefferson drew upon "the harmonizing sentiments of the day, whether expressed in conversation, in letters, printed essays, or in the elementary

peated injuries and usurpations" are explicit objections to: a standing army being kept among the people in times of peace without the consent of local legislatures; the British military operating independent of, and superior to, civilian authorities; and British military members being protected "by a mock Trial, from punishment for any Murders which they should commit on the Inhabitants of [the] States."⁷⁰ Similarly, the popular distrust of centralized control over the sword played a large part in the final ratification of the Articles of Confederation in 1781.⁷¹ Reflecting a conviction that a community of armed citizens was the preferred alternative to meet the security needs of the nation, the Articles heavily favored the creation of state militias in lieu of Congress having the ability to raise a national force in all but the most pressing of circumstances.⁷² Specifically, the Articles prohibited any state from keeping a "vessel of war" or "body of forces" "in time of peace," but required every state to "keep up a well-regulated and disciplined militia, sufficiently armed and accoutered."⁷³ As for the possibility of a national military force, the Articles required that nine states assent to any congressional request to raise "land or sea forces" before those troops could be mustered.⁷⁴

books of public right, as Aristotle, Cicero, Locke, Sidney, etc." MARCUS CUNLIFFE, *SOLDIERS & CIVILIANS, THE MARTIAL SPIRIT IN AMERICA 1775-1865*, at 40 (1968).

⁷⁰ THE DECLARATION OF INDEPENDENCE, *supra* note 69, at para. 17. A final injury and usurpation that focused on the British military was a complaint regarding the "Quartering [of] large bodies of armed troops among [the colonists.]" *Id.* Two years earlier on October 14, 1774, the Continental Congress had adopted a resolution proclaiming "that keeping a standing army in . . . these colonies, in time of peace, without the consent of the legislature of that colony in which such army is kept, is against law." 1 JOURNALS OF THE CONTINENTAL CONGRESS 1774-1789, at 73 (1904).

⁷¹ ARTICLES OF CONFEDERATION (1781), *available at* <http://www.usconstitution.net/articles.html>. See SNOW & DREW, *supra* note 56, at 42. The Continental Congress presented the Articles of Confederation to the states for a vote in November of 1777. BAILYN ET. AL., *supra* note 56, at 281. It took until March of 1781 for all the states to accept the Articles and thereby establish their legality. *Id.*

⁷² ARTICLES OF CONFEDERATION arts. VI, IX.

⁷³ ARTICLES OF CONFEDERATION art. IV.

⁷⁴ ARTICLES OF CONFEDERATION art. IX. "Connecticut proposed an unsuccessful amendment which would have prevented the United States from maintaining an

A series of domestic events over the course of the first half of 1783 affirmed the general distrust of a standing army in time of peace that was explicit in many state constitutions and the Declaration of Independence, and implicit in the Articles of Confederation. During the winter months of that year, a group of disgruntled officers in the Continental Army anonymously circulated through the military camps what came to be known as the Newburgh Addresses.⁷⁵ These documents threatened action by the officers unless Congress met their demands regarding back pay and pension rights that had been promised in 1780.⁷⁶ It ultimately took George Washington to quiet the situation,⁷⁷ and although many heralded the event as evidence of the army rejecting independence from civilian control,⁷⁸ some perceived darker overtones.⁷⁹ The Newburgh Affair was followed shortly by the discovery that Major General Henry Knox and other top army officers had formed the

army in time of peace, even with consent [of the legislature]." CUNLIFFE, *supra* note 69, at 41.

⁷⁵ BAILYN ET. AL., *supra* note 56, at 303; KOHN, EAGLE, *supra* note 24, at 19, 29-31; H. Richard Uviller & William G. Merkel, *Muting the Second Amendment: The Disappearance of the Constitutional Militia*, in *THE SECOND AMENDMENT IN LAW AND HISTORY* 148, 155-56 (Carl T. Bogus ed., 2000). There were two addresses circulated, with the first, distributed on March 10, being the most inciting. KOHN, EAGLE, *supra* note 24, at 29-30. The first address spoke, among other things, of the suffering and glory of the army, "the coldness and severity of government," and the country's ingratitude for the sacrifices made to secure independence. *Id.* at 29.

⁷⁶ Uviller & Merkel, *supra* note 75, at 155-56; BAILYN ET. AL., *supra* note 56, at 303; KOHN, EAGLE, *supra* note 24, at 19, 29-31.

⁷⁷ BAILYN ET. AL., *supra* note 56, at 303. On Saturday, March 15, Washington publicly challenged the motives of those behind the addresses and reminded his officers of their duty and the disgrace that would follow if they should turn their swords against Congress. KOHN, EAGLE, *supra* note 24, at 31-32. This speech came to be known as the "spectacles" speech, because towards its end, Washington, as he put spectacles on in order to read a passage from a letter he held, remarked that he had "not only grown gray, but almost blind in the service of [his] country." Uviller & Merkel, *supra* note 75, at 156; *see also* KOHN, EAGLE, *supra* note 24, at 32.

⁷⁸ *See* KOHN, EAGLE, *supra* note 24, at 34.

⁷⁹ *See* Uviller & Merkel, *supra* note 75, at 156. There is a strong historical record for the conclusion that nationalists in the Congress exploited the discontent of the army for purposes of pressuring the Congress to give the national government the power to tax imports. *See* KOHN, EAGLE, *supra* note 24, at 11-39.

Society of Cincinnati.⁸⁰ Because that group would be secretive and membership would pass to the eldest son in each officer's family, it appeared to many to be a step towards the establishment of a military aristocracy of the kind that grew out of the classic standing army.⁸¹ Finally, in June of 1783, army recruits of the Pennsylvania line barricaded Congress in the State House in Philadelphia, demanding pay and bonuses.⁸² Although the volunteers eventually dispersed, and thus, that crisis, like the others of 1783, passed without civilian government being compromised by the military, the totality of the events reinforced the apprehension that many across the country felt towards the establishment of a national army.⁸³

⁸⁰ See KOHN, EAGLE, *supra* note 24, at 13; Uviller & Merkel, *supra* note 75, at 156.

⁸¹ See BAILYN ET. AL., *supra* note 56, at 289; KOHN, EAGLE, *supra* note 24, at 13; Uviller & Merkel, *supra* note 75, at 156.

⁸² See Uviller & Merkel, *supra* note 75, at 156.

⁸³ In his tenth essay, published on January 24, 1788, the Anti-Federalist essayist "Brutus" used the Newburgh Affair to challenge the military provisions in the proposed Constitution. See *Brutus, Essay of January 24, 1788*, NEW YORK JOURNAL, reprinted in 2 THE COMPLETE ANTI-FEDERALIST 358, 414 (Herbert J. Storing ed., 1981). Recalling this event, "Brutus" wrote:

When an anonymous [sic] writer addressed the officers of the army at the close of the war, advising them not to part with their arms, until justice was done them — the effect it had is well known. It affected them like an electric shock. He wrote like Cesar [sic]; and had the commander in chief, and a few more officers of rank, countenanced the measure, the desperate resolution had been taken, to refuse to disband. What the consequences of such a determination would have been, heaven only knows.— The army were [sic] in the full vigor of health and spirits, in the habit of discipline, and possessed of all our military stores and apparatus. They would have acquired great accessions of strength from the country.— Those who were disgusted at our republican forms of government (for such there then were, of high rank among us) would have lent them all their aid.— We should in all probability have seen a constitution and laws, dictated to us, at the head of an army, and at the point of a bayonet, and the liberties for which we had so severely struggled, snatched from us in a moment. It remains a secret, yet to be revealed, whether this measure was not suggested, or at least countenanced, by some, who have had great influence in producing the present system.— Fortunately indeed for this country, it had at the head of the army, a patriot as well as a general; and many of our principal officers, had not abandoned the characters of citizens, by assuming that of sol-

II. THE FRAMING AND ADOPTION OF A NEW CONSTITUTION

Over the course of the middle 1780s, many came to believe that the military establishment created under the Articles of Confederation was too weak to enable the country to overcome security threats posed by both internal and external forces.⁸⁴ This concern, along with others regarding the structure and power of the national government, led to a congressional resolution calling for delegates from each state to gather in Philadelphia on the second Monday in May of 1787 for the purpose of revising the Articles of Confederation.⁸⁵ Notwithstanding the limited scope of this resolution, the ultimate effect of the process that it instigated was that by December of 1791, the country had an entirely new Constitution and a national Bill of Rights. The place that the military establishment would occupy in the nation's life was a central component of the struggle that surrounded the creation of both documents.

A. *The Constitutional Convention of 1787*

When the Federal Constitutional Convention opened in the late spring of 1787, the ability of the nation to defend itself had the full attention of the convention delegates. Given their old-world frame of reference, many of the delegates accepted that armed aggression was an inevitable part of international life, and they assumed that the newly formed American states, because of their land mass, population, and resources, would be tempting targets for European powers.⁸⁶

diers, and therefore, the scheme proved abortive. But are we to expect, that this will always be the case?

Id. (footnote omitted).

⁸⁴ See *infra* notes 103-10 and accompanying text.

⁸⁵ See 1 DOCUMENTARY HISTORY OF THE CONSTITUTION OF THE UNITED STATES 1786-1870, at 8 (1998).

⁸⁶ Richard H. Kohn, *The Constitution and National Security: The Intent of the Framers*, in THE UNITED STATES MILITARY UNDER THE CONSTITUTION OF THE UNITED STATES, 1789-1989, 61, at 64 (Richard H. Kohn ed., 1991); see also 1 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, at 466 (Max Farrand rev. ed., 1966) [hereinafter RECORDS]. Alexander Hamilton warned his fellow delegates on

Beyond this general foreboding of an external threat, many of the delegates also feared three specific and timely outside dangers. First, the British, who were still entrenched in their Canadian territories, represented a real and formidable adversary.⁸⁷ Notwithstanding the outcome of the Revolutionary War, the British had not given up hope of reclaiming their lost American colonies and, in any event, were eager to thwart America's western imperial dreams.⁸⁸ Second, the Spanish lay close by in Florida and Louisiana, and Spain feared that American expansion into the Mississippi Valley would weaken its hold over its colonial possessions.⁸⁹ As did Britain's continuing occupation of strategically located forts on American soil after the Revolutionary War,⁹⁰ Spain's closing of a portion of the Mississippi River to American commerce in 1784 signaled its intention to deal aggressively with the new American states.⁹¹ Third, delegates to the Constitutional Convention saw the Native Americans as a significant external threat

June 29 that a dissolution of the Union, followed by the establishment of partial confederacies, would inevitably lead to alliances being formed "with different rival & hostile nations of Europes [sic], who will foment disturbances among ourselves, and make us parties to all their own quarrels." *Id.*

⁸⁷ Kohn, *supra* note 86, at 65; KOHN, EAGLE, *supra* note 24, at 288; FRANK MONAGHAN, JOHN JAY 361 (1935).

⁸⁸ Kohn, *supra* note 86, at 65.

⁸⁹ *Id.*

⁹⁰ See KOHN, EAGLE, *supra* note 24, at 51, 66; MONAGHAN, *supra* note 87, at 361. The British would not relinquish all of their military posts in U.S. territory until after it signed Jay's Treaty in 1794. See BAILYN ET. AL., *supra* note 56, at 326; MONAGHAN, *supra* note 87, at 381. Article 2 of Jay's Treaty called for the British to "withdraw all . . . troops and garrisons from all posts and places within the boundary lines assigned by the treaty of peace to the United States." *Treaty of Amity, Commerce, and Navigation, between His Britannic Majesty, and the United States of America, by Their President, with the Advice and Consent of Their Senate*, in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 520, 520 (Walter Lowrie & Walter S. Franklin eds., 1832). Britain justified its refusal to abandon the forts earlier by claiming that the United States had not honored its commitments to recommend to the states that loyalist property confiscated during the Revolution be restored and not to make laws obstructing the recovery of prewar debts. BAILYN ET. AL., *supra* note 56, at 304.

⁹¹ BAILYN ET. AL., *supra* note 56, at 305; JON KUKLA, A WILDERNESS SO IM-MENSE: THE LOUISIANA PURCHASE AND THE DESTINY OF AMERICA 59, 113-14 (2003).

to the country's security.⁹² The frontiers of the new states were frequently raided by Native Americans acting on their own or as a result of encouragement and support from Britain and Spain, which were both eager to use the natives as allies in the countries' efforts to turn back America's advance to the west.⁹³

In addition to these external dangers, Federal Convention delegates were also concerned about the nation being undermined from within. Like the case with the outside forces, the internal hazards took the form of a general foreboding and a specific menace. As for the former, delegates saw the weak central government created by the Articles of Confederation as giving rise to a collection of largely autonomous states that were in danger of drifting apart and becoming rivals.⁹⁴ In the words of John Jay, written after the close of the Federal Convention, many of the delegates saw the American states in the mid-1780s as "[a] prey to discord, jealousy, and mutual injuries . . . formidable only to each other."⁹⁵

The specific internal danger on the minds of the delegates was the possibility of armed insurrection, as manifested just a

⁹² See KOHN, EAGLE, *supra* note 24, at 92-95.

⁹³ See BAILYN ET. AL., *supra* note 56, at 321; KOHN, EAGLE, *supra* note 24, at 65, 150.

⁹⁴ KOHN, EAGLE, *supra* note 24, at 64. James Madison, on June 29th, spoke to this danger at the convention saying:

The same causes which have rendered the old world the Theatre of incessant wars, & have banished liberty from the face of it, wd. [sic] soon produce the same effects here. The weakness & jealousy of the small States wd. [sic] quickly introduce some regular military force agst. [sic] sudden danger from their powerful neighbours [sic]. The example wd. [sic] be followed by others, and wd. [sic] soon become universal.

1 RECORDS, *supra* note 86, at 464-65.

⁹⁵ THE FEDERALIST NO. 5, at 30 (John Jay) (E.H. Scott ed., 1898) (alteration omitted). Benjamin Franklin used similarly colorful language at the Constitutional Convention on September 17th to describe the danger presented by disunity among the states when he referred to "enemies" who were "waiting with confidence to hear that our councils are confounded like those of the Builders of Babel; and that our States are on the point of separation, only to meet hereafter for the purpose of cutting one another's throats." 2 RECORDS, *supra* note 86, at 642-43.

few months earlier by Shays' Rebellion.⁹⁶ Stretching over the last few months of 1786 and the early months of 1787, an armed uprising of nearly 2000 farmers in western Massachusetts had taken place in response to what was believed to be unjust actions by, and on behalf of, the mercantile elite of the eastern portion of that state.⁹⁷ This uprising included the occupation of the courts in order to prevent the imprisonment of debtors and an attack on the federal arsenal in Springfield on January 25, 1787.⁹⁸ It ultimately took over 4000 militiamen to put the revolt down.⁹⁹ Shays' Rebellion and the general threat of discord among the confederated states weighed heavily on the minds of delegates as they met in Philadelphia to reconsider the form of the national government.¹⁰⁰

⁹⁶ See Paul Finkelman, "A Well Regulated Militia:" *The Second Amendment in Historical Perspective*, in *THE SECOND AMENDMENT IN LAW AND HISTORY* 117, 117 (Carl T. Bogus ed., 2000).

⁹⁷ BAILYN ET. AL., *supra* note 56, at 306; DAVID P. SZATMARY, SHAYS' REBELLION: THE MAKING OF AN AGRARIAN INSURRECTION 19-36 (1980); see also KOHN, EAGLE, *supra* note 24, at 74 ("Long burdened by heavy debts, mounting taxation, and a lack of hard currency, farmers and townspeople west of Worcester turned on the courts and the legal system with mob action and violence.").

⁹⁸ BAILYN ET. AL., *supra* note 56, at 306; SZATMARY, *supra* note 97, at 58-59, 101-02.

⁹⁹ See SZATMARY, *supra* note 97, at 100, 104-05. The Springfield Arsenal was defended by General William Shepard with 1000 local militiamen. *Id.* at 100-01. Over 3000 militiamen under the command of General Benjamin Lincoln finally defeated the insurgents at Petersham on February 4th, 1787. *Id.* at 104-05.

¹⁰⁰ See *id.* at 67; DAVID C. WILLIAMS, *THE MYTHIC MEANINGS OF THE SECOND AMENDMENT* 42 (2003); see also Levin, *supra* note 20, at 1034 (noting that Secretary of War Henry Knox's warnings of anarchy and imminent civil war after traveling in his home state of Massachusetts "prompted Washington and others to alarmed concern about the need for a national military force to combat such popular upheaval"). But see generally Letter from Thomas Jefferson, Paris, to James Madison (Jan. 30, 1787), in *THOMAS JEFFERSON SELECTED WRITINGS* 67, 68 (Harvey C. Mansfield, Jr., ed., 1979):

I hold it that a little rebellion now and then is a good thing, and as necessary in the political world as storms in the physical. Unsuccessful rebellions, indeed, generally establish the encroachments on the rights of the people which have produced them. An observation of this truth should render honest republican governors so mild in their punishment of rebellions as not to discourage them too much. It is a medicine necessary for the sound health of government.

Id.

A large number of the delegates to the Constitutional Convention had been members of the wartime Colonial Congress or of a wartime state government, or were veterans of the militia or Continental Army, and thus, they brought with them vivid memories of the difficulties that had been experienced in defeating the British.¹⁰¹ Chief among those memories were those that recalled problems in raising the money, arms, and men necessary to power the revolution.¹⁰² Drawing upon this collective experience, many of the delegates believed it was imperative to locate the power to defend the nation in a strong central government, and this naturally led them to be highly critical of the Articles of Confederation.¹⁰³

Under the Articles of Confederation, the central government was relegated to having to come "hat-in-hand" to ask the states for men and money in the event of a need to raise troops.¹⁰⁴ As mentioned earlier, once the Congress had asked for land or sea forces, nine states out the thirteen would have to be present and agree with that request before the support would be given.¹⁰⁵ This difficulty in securing collective action meant that a state might have to go it alone when meeting a threat, and this raised great concerns with many of the delegates. These men were worried about what they perceived to be a collective burden falling unevenly on individual states, some of which they feared were not strong enough to see after their own welfare and others of which were perceived to be indispensable to the security of the entire nation, and thus, could not be allowed to fall if the country was to survive.¹⁰⁶

¹⁰¹ See Kohn, *supra* note 86, at 68; see also BAILYN ET. AL., *supra* note 56, at 307-08 ("Thirty-nine [of the delegates] had served in Congress at one time or another, 8 had worked in the state constitutional conventions, 7 had been state governors, and. . . [o]ne-third were veterans of the Continental Army . . .").

¹⁰² Kohn, *supra* note 86, at 68; Levin, *supra* note 20, at 1034.

¹⁰³ See Kohn, *supra* note 86, at 70; Levin, *supra* note 20, at 1034.

¹⁰⁴ See Kohn, *supra* note 86, at 69.

¹⁰⁵ See *supra* note 74 and accompanying text; see also Kohn, *supra* note 86, at 69 ("[I]f a state was represented by only two delegates, as was often the case, and one did not vote in the affirmative, then that state could not be counted as consenting to the measure.").

¹⁰⁶ See Kohn, *supra* note 86, at 70. The prime example of such a state was

The answer to all of the above concerns for many of the Constitutional Convention delegates was for the central government to be empowered to form and pay for a standing army that could prepare in peace to meet the challenges of war and insurrection.¹⁰⁷ Consistent with that position, in his address to the convention on May 29, 1787, which opened the main business of that body, Virginia Governor Edmund Randolph stated that any government that resulted from the delegates' efforts must have the character to be able "to secure . . . against foreign invasion [and] . . . dissensions between members of the Union, or seditions in particular states."¹⁰⁸ Although he praised the authors of the Articles of Confederation as "patriots," Governor Randolph nevertheless claimed those Articles were defective on both counts.¹⁰⁹ Specifically, Governor Randolph said that the Articles did not permit Congress "to prevent a war nor to support it by th[eir] own authority" or to "check thè quarrrals [sic] between states" or put down a rebellion in one of the same.¹¹⁰

New York, which represented a critical strategic asset because of its protected harbor, important navigable river and physical location adjacent to New England, Canada, and the middle states. *See id.* at 71.

¹⁰⁷ *See id.* at 71 ("The framers believed not only that the authority to prepare had to be spelled out clearly but also that some establishment in peacetime was needed regardless of the threat, merely to develop in the nation the expertise of military leadership and to keep alive the technical knowledge for combat."); *see also* 2 RECORDS, *supra* note 86, at 330 ("Mr. Dayton. [P]reparations for war are generally made in peace; and a standing force of some sort may, for ought we know, become unavoidable.").

¹⁰⁸ RECORDS, *supra* note 86, at 18. In a letter written in October of 1787, Governor Randolph echoed his statements before the convention when he wrote that the central government must be empowered to serve as "[a] shield against foreign hostility, and a firm resort against domestic commotion." Kohn, *supra* note 86, at 70.

¹⁰⁹ RECORDS, *supra* note 86, at 18-19.

¹¹⁰ *Id.* at 19 (alteration in original). Pierce Butler, a delegate from South Carolina, echoed Governor Randolph's remarks on June 11 when he declared that "money is strength" and then immediately offered up a boyhood memory of having read that Julius Caesar declared that "if he had but money he would find soldiers, and every thing necessary to carry on a war." *Id.* at 204. Throughout the Constitutional Convention, Governor Randolph fluctuated between the Federalist and Anti-Federalist points of view. *See America's Founding Fathers: Delegates to the Constitutional Convention*, The National Archives Experience, at

It was Alexander Hamilton, serving as a New York delegate to the convention, who took Governor Randolph's general pronouncements and put them into a most emphatic form.¹¹¹ On June 18, Hamilton rose to speak for the first time at the convention to share his vision of a properly formed "federal government."¹¹² Calling attention to the internal situation in Massachusetts and the external threat posed by "[f]oreign powers," Hamilton claimed that "[a] certain portion of military force is absolutely necessary in large communities."¹¹³ While citing, among other examples, the wars between the Swiss cantons, Hamilton argued that because the Articles of Confederation placed the security power in the hands of the state governments, the concept of the nation as a confederacy was rendered an unlikely proposition.¹¹⁴ The cure for this ill, in Hamilton's mind, was to vest the war powers in a "supreme Executive authority of the United States" and a national senate.¹¹⁵ As for the question of these powers remaining with

http://www.archives.gov/national-archives-experience/charters/constitution_founding_fathers.html (last visited Oct. 4, 2006). He ultimately refused to sign the Constitution at the end of the Federal Convention because he felt that it was not sufficiently republican. *See id.* However, by the time of the Virginia convention for ratification, Governor Randolph supported the Constitution and worked to win its approval. *See id.* The reason behind his change of heart was that eight states had already ratified the document by the time of the Virginia convention, so he saw the question of being simply "Union or no Union." *See id.*

¹¹¹ Governor Randolph's "Virginia Plan," which was put forth on May 29, apparently assumed a troop-raising power even though it did not specify one because it gave the "National Legislature" the power "to call forth the force of the union agst. [sic] any member of the Union failing to fulfill its duty under the articles thereof." 1 RECORDS, *supra* note 86, at 21.

¹¹² *Id.* at 282-83. "Mr. Hamilton had been hitherto silent on the business before the Convention, partly from respect to others whose superior abilities age & experience rendered him unwilling to bring forward ideas dissimilar to theirs, and partly from his delicate situation with respect to his own State, to whose sentiments as expressed by his Colleagues, he could by no means accede." *Id.* at 282.

¹¹³ *Id.* at 285. Of the foreign powers, Mr. Hamilton said, "[t]hey will interpose, the confusion will increase, and a dissolution of the Union ensue." *Id.*

¹¹⁴ *See id.* at 285-86. The Swiss Confederation would later be used by the Anti-Federalists as a model of government that allowed its citizens to "retain[] their independence, republican simplicity and valour." 1 Storing, *supra* note 83, at 20 (citation omitted).

¹¹⁵ *See* 1 RECORDS, *supra* note 86, at 292. The Senate and another body called

the states, Hamilton put forth the proposition that "[n]o State . . . have any forces land or Naval; and the Militia of all the States . . . be under the sole and exclusive direction of the United States, the officers of which to be appointed and commissioned by them."¹¹⁶

On the opposite side of the debate from Hamilton were delegates to the Federal Convention for whom the standing army controversy became part of their motivation to oppose the ratification of the entire Constitution. Chief among these critics was Elbridge Gerry, a delegate from Massachusetts who had long been a leading voice against a peacetime military establishment.¹¹⁷ Following a motion on August 18 from Nathaniel Gorham, also of Massachusetts, to modify a proposed clause from "[T]o raise armies" into "[T]o raise and support armies," Gerry proclaimed that the absence of a check against standing armies in time of peace would generate great public opposition to the clause.¹¹⁸ Saying that he could never consent to a power to keep up an indefinite number of troops in time of peace, Gerry proposed that a limit of two or three thousand troops be attached to the clause, but the wider body disagreed.¹¹⁹ Gerry returned to the theme of standing armies on September 5 when the convention again took up the "raise and support armies" clause.¹²⁰ Similar to his earlier com-

the "Assembly" would make up "the Legislature of the United States." *Id.* at 291.

¹¹⁶ *Id.* at 293. Hamilton had long believed that the country needed a centrally managed military force. See KOHN, EAGLE, *supra* note 24, at 47-48. A congressional committee under his leadership had proposed in 1783 that the country, in place of relying upon the militia, form a 3000 man national army and create an elite corps of volunteer citizens. *Id.*

¹¹⁷ KOHN, EAGLE, *supra* note 24, at 52-53, 61; see also *America's Founding Fathers: Delegates to the Constitutional Convention*, The National Archives Experience, at http://www.archives.gov/national-archives-experience/charters/constitution_founding_fathers.html (last visited Oct. 4, 2006). Gerry was at first an advocate of a strong central government, but he ultimately refused to sign the Constitution because it lacked a bill of rights, and he deemed it a threat to republicanism. *Id.*

¹¹⁸ 2 RECORDS, *supra* note 86, at 329.

¹¹⁹ See *id.* at 329-30. Gerry's formal motion to put a cap on the number of troops that could be raised in time of peace was joined by Luther Martin, a delegate from Maryland. *Id.* at 330.

¹²⁰ See *id.* at 509. The delegates were considering a recommendation from the

ments on the matter, Gerry condemned a standing army as being "dangerous to liberty" and again argued that if one was necessary, some restriction of its number and duration should be in the Constitution.¹²¹ As before, Gerry's efforts were met with no success.¹²²

In the end, a consensus arose at the Federal Convention in favor of trying to find a middle way that would strike a balance between the country's security needs and the dangers presented by a standing army. One of the leading spokesmen for this compromise approach was James Madison of Virginia. On August 23, when the convention debated the status of the militia under the new constitution, Madison said, "[a]s the greatest danger is that of disunion of the States, it is necessary to guard agst. [sic] it by sufficient powers to the Common Govt. [sic] and as the greatest danger to liberty is from large standing armies, it is best to prevent them by an effectual provision for a good Militia."¹²³ Thereafter, on September 14,

Committee of Eleven to add to the clause "to raise and support armies" the words "but no appropriation of money to that use shall be for a longer term than two years." See *id.* at 508.

¹²¹ *Id.* at 509. Roger Sherman, a delegate from Connecticut, remarked that he too would "like a reasonable restriction on the number and continuance of an army in time of peace." *Id.* Sherman, who was instrumental in Connecticut's ratification of the Constitution, was a prime mover behind the Great Compromise, which broke the deadlock between the large and small states over representation. See *America's Founding Fathers: Delegates to the Constitutional Convention*, The National Archives Experience, at http://www.archives.gov/national-archives-experience/charters/constitution_founding_fathers.html (last visited Oct. 4, 2006).

¹²² See 2 RECORDS, *supra* note 86, at 509.

¹²³ See *id.* at 388. Mr. Madison appeared to believe that only an effective militia under the auspices of the national government could thwart the move towards the creation of a national standing army to meet the country's security needs. See Levin, *supra* note 20, at 1040. Mr. Madison insisted that

[i]f insurrections should arise, or invasions should take place, the people ought unquestionably to be employed, to suppress and repel them, rather than a standing army. The best way to do these things was to put the militia on a good and sure footing, and enable the government to make use of their services when necessary.

3 ELLIOT'S DEBATES ON THE FEDERAL CONSTITUTION, 378 (2d ed. 1891). See generally 2 RECORDS, *supra* note 86, at 387 ("The States neglect their Militia now, and the more they are consolidated into one nation, the less each will rely on its own interior provisions for its safety & the less prepare its Militia for that pur-

when Madison's fellow Virginian, George Mason,¹²⁴ moved to add the words, "[a]nd that the liberties of the people may be better secured against the danger of standing armies in time of peace" to the beginning of a clause that concerned the militia,¹²⁵ Madison availed himself of the opportunity once again to state the case for a middle way. Consistent with the view he expressed on August 23, Madison voiced his support of Mason's motion by saying that "as armies in time of peace are allowed on all hands to be an evil, it is well to discountenance them by the Constitution, as far as will consist with the essential power of the Govt. [sic] on that head."¹²⁶ Governor Morris of Pennsylvania opposed the motion on the grounds that it would "set[] a dishonorable mark of distinction on the military class of Citizens."¹²⁷ The motion failed two votes to nine.¹²⁸

pose. . . . The Discipline of the Militia is evidently a *National* concern, and ought to be provided for in the *National* Constitution.").

¹²⁴ George Mason would eventually leave the convention without signing the new Constitution. See MALCOLM, *supra* note 37, at 154. Among other reasons given for his displeasure with the final product, Mason decried the absence of a Declaration of Rights in the Constitution, and he believed that the Constitution created a House of Representatives that was not truly representative of the nation, a Senate that was too powerful, and a national judiciary that would destroy the state judiciaries. See *id.* at 154-55; 2 RECORDS, *supra* note 86, at 637-38.

¹²⁵ 2 RECORDS, *supra* note 86, at 617. Governor Randolph, in an Anti-Federalist moment, seconded Colonel Mason's motion. 2 RECORDS, *supra* note 86, at 617. See *supra* note 110 and accompanying text.

The clause in question was ultimately placed in Article I to the Constitution. See U.S. CONST. art. I, § 8, cl. 16 ("The Congress shall have Power . . . To provide for organizing, arming, and disciplining, the Militia, and for governing such Part of them as may be employed in the Service of the United States, reserving to the States respectively, the Appointment of the Officers, and the Authority of training the Militia according to the discipline prescribed by Congress").

¹²⁶ 2 RECORDS, *supra* note 86, at 617. Charles Pinckney of South Carolina had earlier offered his view of a middle ground in a document he submitted to the Constitutional Convention on May 28, titled *Observations on the Plan of Government*. 3 RECORDS, *supra* note 86, at 106. Mr. Pinckney argued that standing armies were "contrary to . . . the nature of . . . Government," and thus, the only place the country could turn in the case of necessity was the militia. *Id.* at 118. In light of this, Mr. Pinckney argued that the exclusive right of establishing regulations for the militia ought to lie with the national government so as to ensure uniformity in discipline and military habits. *Id.*

¹²⁷ 2 RECORDS, *supra* note 86, at 617.

¹²⁸ *Id.*

It becomes apparent that one of the most hotly contested issues within the context of the standing army debate at the Federal Convention was the status and form that the militia would take under the new Constitution.¹²⁹ For many within and outside the convention, the militia served as both "metaphor and actuality: metaphor for the self-governing community of citizens and actuality of the active practice of that citizenship to guard against the corruption of the central state."¹³⁰ More pointedly, the militia was widely seen as representing "the very antithesis of the corrupting and tyrannical standing army."¹³¹

In a more practical vein, many believed that a citizen militia was a force that could meet the security needs of the country.¹³² At the Federal Convention on June 4, George Mason spoke of the power presented by a group of citizens in their love, affection, and attachment "to their laws, to their freedom, and to their country."¹³³ Mason said of the militia that "[e]very husbandman will be quickly converted into a soldier when he knows and feels that he is to fight not in defence [sic] of the rights of a particular family, or a prince, but for his own."¹³⁴ Moreover, the militia's supporters were able to add a measure of weight to their argument by pointing to some valuable contributions made by that body to the suc-

¹²⁹ See Levin, *supra* note 20, at 1034-35.

¹³⁰ *Id.* at 1040 (footnote omitted). In debating a provision to grant to the national government the power organize, arm, and discipline the militia, Elbridge Gerry posed the rhetorical question, "Will any man say that liberty will be as safe in the hands of eighty or a hundred men taken from the whole continent, as in the hands of two or three hundred taken from a single State?" See 2 RECORDS, *supra* note 86, at 386.

¹³¹ KOHN, EAGLE, *supra* note 24, at 8.

¹³² See, e.g., SNOW & DREW, *supra* note 56, at 31 ("Colonial success [in the Revolutionary War] created the myth that militia was effective against regular troops and that the United States did not need a standing armed force of any size during peacetime. The militia could be quickly mobilized and hold its own until a regular army could be fielded to carry the day.").

¹³³ 1 RECORDS, *supra* note 86, at 112.

¹³⁴ *Id.* Mr. Mason went on to cite the Grecian republics, the States of Holland, the Swiss Cantons, and the American experience in the Revolutionary War as examples that proved the rule he was advancing. *Id.* at 112-13.

cess of the Revolutionary War. For example, during that struggle the militia "battled the Indians and Tories along the Appalachian frontier, blunted Royal Naval raids from the sea, carried on partisan warfare against the British regulars and Tories in the mid-Atlantic states and the South, and enforced the laws passed by the Whig assemblies."¹³⁵ There were many others, however, who held less flattering opinions of the militia.

For a large number of delegates at the Constitutional Convention, the militia had fundamental weaknesses that rendered it a less than optimal fighting force.¹³⁶ Many of the delegates would no doubt have been aware that George Washington, towards the end of the campaign for New York City and northern New Jersey in September of 1776, had said of the militia, "[t]o place any dependence upon [it] is, assuredly, resting upon a broken staff."¹³⁷ Perhaps inspired by General

¹³⁵ Allan R. Millett, *The Constitution and the Citizen-Soldier*, in *THE UNITED STATES MILITARY UNDER THE CONSTITUTION OF THE UNITED STATES, 1789-1989*, 97, 100 (Richard H. Kohn ed., 1991); see also SNOW & DREW, *supra* note 56, at 41-42, 54 (discussing the role of the militia in suppressing Tory support and harassing British regulars, and the defeat of Hessian mercenaries in Bennington, Vermont, by militia).

¹³⁶ See Millett, *supra* note 135, at 100-01; see also ROBERT J. SPITZER, *THE RIGHT TO BEAR ARMS* 18 (2001). On the performance of the militia during the Revolutionary War, Mr. Spitzer writes:

[The militia] proved to be an administrative and military nightmare. State detachments could not be easily combined into larger fighting units; soldiers could not be relied on to serve for extended periods, and desertions were common; militia detachments would often enter the army encampment, consume food and supplies, and then leave before battle; officers were elected based on popularity rather than experience or training; discipline and uniformity were almost nonexistent.

Id.

¹³⁷ Letter From George Washington to The President of Congress (Sept. 24, 1776), reprinted in *AMERICAN MILITARY THOUGHT* 9, 12 (Walter Millis ed., 1966). General Washington continued:

Men just dragged from the tender Scenes of domestick [sic] life; unaccustomed to the din of Arms; totally unacquainted with every kind of Military skill, which being followed by a want of confidence in themselves, when opposed to Troops regularly train'd [sic], disciplined, and appointed, superior in knowledge, and superior in Arms, makes them timid, and ready to fly from their own shadows. Besides, the sudden change in

Washington's statement and their own experience of the militia's limitations,¹³⁸ some of the delegates preferred to do away with state centered militias entirely. Charles Pinckney, a delegate from South Carolina, submitted a document titled "*Observations on the Plan of Government*" on May 28 to the Federal Convention, which called for the militia to be, "as far as possible," a national force subject to "uniformity in Discipline and Regulations."¹³⁹ Thereafter, Alexander Hamilton's proposal of June 18 would have accomplished the result advocated by Pinckney by putting the militia "under the sole and exclusive direction of the United States."¹⁴⁰ Similarly, on August 18, General Charles Cotesworth Pinckney, also of South Carolina, argued that the dissimilarities between the state militias produced "the most serious mischiefs," and he complained that if left to their own devices "[t]he States would never keep up a proper discipline of their militia."¹⁴¹ Later in the midst of the same debate that produced the preceding comments, General Pinckney, declaring that he had "but a scanty faith in the Militia," called for "a real military force" to

their manner of living, (particularly in the lodging) brings on sickness in many; impatience in all, and such an unconquerable desire of returning to their respective homes that it not only produces shameful, and scandalous Desertions among themselves, but infuses the like spirit in others.

Id.

¹³⁸ See KOHN, EAGLE, *supra* note 24, at 9-10; see also Millett, *supra* note 135, at 100 (explaining that during the colonial period, the militias "could not be used for operations that required troops to cross colonial boundaries, and it could not be deployed as units beyond limited compulsory periods of service").

¹³⁹ 3 RECORDS, *supra* note 86, at 118. Mr. Pinckney continued:

To place therefore a necessary and Constitutional power of defence [sic] and coercion in the hands of the Federal authority, and to render our Militia uniform and national, I am decidedly in opinion they should have the exclusive right of establishing regulations for their Government and Discipline, which the States should be bound to comply with, as well as with their Requisitions for any number of Militia, whose march into another State, the Public safety or benefit should require.

Id. at 119.

¹⁴⁰ RECORDS, *supra* note 86, at 293. See *supra* note 116 and accompanying text.

¹⁴¹ RECORDS, *supra* note 86, at 330.

be created.¹⁴² However, notwithstanding these petitions for a national force in the place of the state centered militias, most delegates to the Constitutional Convention recognized that the latter were a political reality that had to be taken into account.¹⁴³

The combination of the belief that standing armies were a threat to liberty and concerns about the cost of maintaining a large permanent force led to the conclusion that the nation would depend upon the militia, under some amount of state control, to be a significant part of the country's military establishment.¹⁴⁴ A compromise on the status of the militia, therefore, became a primary feature of the effort to strike a balance between the country's security needs and the dangers presented by a standing army.¹⁴⁵ Consistent with the middle way

¹⁴² *Id.* at 332. Between these comments and those recounted in the previous sentence, General Pinckney had argued that to split the regulation and discipline of militia between the general government and the state governments "w[oul]d be an incurable evil." *Id.* at 331. Pierce Butler of South Carolina joined General Pinckney and "urged the necessity of submitting the whole Militia to the general Authority, which had the care of the general defence [sic]." *Id.* Recalling the debates in the Constitutional Convention over the status of the militia, Governor Morris said in 1815 that the chief danger that he and like-minded others sought to avoid "was the hazarding of the national safety by a reliance on that expensive and inefficient force." KOHN, *EAGLE*, *supra* note 24, at 283 (quoting RECORDS, *supra* note 86, at 420). Nevertheless, Elbridge Gerry insisted that any plan that called for the militia to be put under exclusive national control would "have as black a mark as was set on Cain." 2 RECORDS, *supra* note 86, at 332.

¹⁴³ See MALCOLM, *supra* note 37, at 153; Millett, *supra* note 135, at 101.

¹⁴⁴ SPITZER, *supra* note 136, at 21. Compelled by the political realities of the times, even George Washington supported the continuation of the state militias; however, he argued for the creation of a select militia that would be formed from the most able militiamen to serve as the foundation of the national defense. General George Washington, *Sentiments on a Peace Establishment*, in AMERICAN MILITARY THOUGHT, *supra* note 137, at 24. See RUSSELL F. WEIGLEY, *TOWARDS AN AMERICAN ARMY* 10-13 (1962).

¹⁴⁵ Other provisions in the Constitution that reflect the compromise that was struck over the issue of standing armies are those that split authority over the military forces between the Congress and the President. Compare U.S. CONST. art. I, § 8, cls. 11-16 (detailing Congress's powers over the country's military establishment), with *id.* at art. 2, § 2, cl. 1 (the President as Commander in Chief of the armed forces), and *id.* at art. I, § 8, cl. 12 (the restriction on any congressional appropriation of money for the armed forces to a term of no more than two years).

advocated by James Madison,¹⁴⁶ the Constitution that was sent to the states for ratification gave the central government more control over the militia than it had enjoyed under the Articles of Confederation, but that control was not exclusive.¹⁴⁷ While the new Constitution explicitly gave Congress the power to organize, arm, and discipline the militia,¹⁴⁸ call the militia into federal service,¹⁴⁹ and govern that body once it was "employed in the Service of the United States,"¹⁵⁰ it reserved to the states the authority to appoint officers to the militia and to train that body according to the discipline prescribed by Congress.¹⁵¹

B. The Ratification Debate

Once the Federal Convention adjourned, the battle over the Constitution moved into the ratification process and here, as had been the case in Philadelphia, the dangers presented

¹⁴⁶ See *supra* notes 123-28 and accompanying text.

¹⁴⁷ See Kohn, *supra* note 86, at 76 (The compromise reached was to give the Congress and the states concurrent power over the militia "similar to the concurrent taxing power").

¹⁴⁸ 2 RECORDS, *supra* note 86, at 656; see also U.S. CONST. art. I, § 8, cl. 16 ("The Congress shall have Power . . . To provide for organizing, arming, and disciplining, the Militia, and for governing such Part of them as may be employed in the Service of the United States, reserving to the States respectively, the Appointment of the Officers, and the Authority of training the Militia according to the discipline prescribed by Congress").

¹⁴⁹ 2 RECORDS, *supra* note 86, at 656; see also U.S. CONST. art. I, § 8, cl. 15 ("The Congress shall have Power . . . To provide for calling forth the Militia to execute the Laws of the Union, suppress Insurrections and repel Invasions").

¹⁵⁰ RECORDS, *supra* note 86, at 656; see also U.S. CONST. art. I, § 8, cl. 16. The Constitution that was sent forward for ratification also designated the President as the commander-in-chief of the militia when called into federal service. 2 RECORDS, *supra* note 86, at 659; see also U.S. CONST. art. II, § 2, cl. 1.

¹⁵¹ RECORDS, *supra* note 86, at 656; see also U.S. CONST. art. I, § 8, cl. 16. James Madison had moved for the states to have the power to appoint only those officers "under the rank of General officers," but this was met with stiff opposition. See 2 RECORDS, *supra* note 86, at 388 (emphasis omitted). Roger Sherman of Connecticut considered Mr. Madison's proposal "absolutely inadmissible" and said "that if the people should be so far asleep as to allow the Most influential officers of the Militia to be appointed by the Genl. [sic] Government, every man of discernment would rouse them by sounding the alarm to them." *Id.* Mr. Madison's proposal was defeated eight votes to three. *Id.*

by a national standing army were a central part of the debate.¹⁵² Luther Martin, who left the Federal Convention without signing the Constitution, gave a speech to the Maryland legislature in November of 1787 in which he attacked what he perceived to be the proposed power of Congress to establish a national standing army.¹⁵³ Martin asserted that the plan of government contained in the new Constitution, "instead of guarding against a standing army, that engine of arbitrary power . . . has in its formation given it an express and constitutional sanction."¹⁵⁴ In Pennsylvania, a minority of the delegates to that state's ratifying convention proposed that the Constitution be amended before adoption so as to, among other things, prohibit the raising of a standing army in time of peace, require that the military "be kept under strict subordination to and be governed by the civil powers," and leave to the individual states the power of "organizing, arming and disciplining the militia."¹⁵⁵ In June of 1788, at the Virginia

¹⁵² See MALCOLM, *supra* note 37, at 155; Levin, *supra* note 20, at 1043-44.

¹⁵³ 3 RECORDS, *supra* note 86, at 207-09.

¹⁵⁴ *Id.* at 207 (emphasis omitted). Maryland ultimately voted to ratify the Constitution. See 1 ELLIOT'S *supra* note 123, at 324-25. However, a minority of delegates to the Maryland ratifying convention sought to protect against the dangers of standing armies by proposing amendments that would require a two-thirds majority in Congress to raise troops in time of peace and protect the rights of conscientious objectors. BILL OF RIGHTS, *supra* note 67, at 181.

¹⁵⁵ *The Address and Reasons of Dissent of the Minority of the Convention of Pennsylvania to Their Constituents*, PENNSYLVANIA PACKET AND DAILY ADVISOR, Dec. 18, 1787, reprinted in 3 Storing, *supra* note 83, at 151-52. In the essay, *Address and Reasons of the Dissent of the Minority of the Convention of Pennsylvania To Their Constituents*, published on December 17, 1787, in the *Pennsylvania Packet and Daily Advertiser*, the minority delegates complained further along these lines saying:

A standing army in the hands of government placed so independent of the people, may be made a fatal instrument to overturn the public liberties; it may be employed to enforce the collection of the most oppressive taxes, and to carry into execution the most arbitrary measures. An ambitious man who may have the army at his devotion, may step up into the throne, and seize upon absolute power.

Id. at 164; see also BILL OF RIGHTS, *supra* note 67, at 181 (describing a minority at the Massachusetts ratifying convention who proposed that the Constitution should never be construed to authorize Congress "to raise standing armies, unless when necessary for the defence [sic] of the United States, or of some one or more

ratifying convention in Richmond, Patrick Henry kept up the drumbeat against a national standing army by calling Congress's proposed ability to raise and support an army "a most dangerous power," which, if left unbounded, would "lead to despotism."¹⁵⁶ Giving heed to Henry's words, the Virginia convention, after ratifying the Constitution, proclaimed that "standing armies in time of peace are dangerous to liberty" and proposed amendments to the Constitution that would: recognize the right of the people to bear arms and the importance of a well regulated militia;¹⁵⁷ require that "the military . . . be under strict subordination to and governed by the Civil power;"¹⁵⁸ restrict the ability of the national government to quarter soldiers in private homes;¹⁵⁹ and require a two-thirds vote of Congress to raise and keep standing armies in peacetime.¹⁶⁰ Echoing Virginia, the state conventions in New Hampshire,¹⁶¹ New York,¹⁶² North Carolina,¹⁶³ and

of them").

¹⁵⁶ 3 ELLIOT'S, *supra* note 123, at 410. Henry also feared that the weakening of the states' control over their militias under the Constitution would leave the states and the people vulnerable to a national standing army. *See id.* at 385; *see also id.* at 379 (reiterated the fact that George Mason declared before the Virginia ratifying convention on June 14, 1788, that "[he] abominate[d] and detest[ed] the idea of a government, where there is a standing army").

¹⁵⁷ BILL OF RIGHTS, *supra* note 67, at 182.

¹⁵⁸ *Id.* at 182-83 (prefacing this demand with the assertion that "standing armies in time of peace are dangerous to liberty").

¹⁵⁹ *Id.* at 216 ("Eighteenth. That no Soldier in time of peace ought to be quartered in any house without the consent of the owner, and in time of war in such manner only as the laws direct.").

¹⁶⁰ The Avalon Project at Yale Law School, *Ratification of the Constitution by the State of Virginia; June 26, 1788*, <http://www.yale.edu/lawweb/avalon/const/ratva.htm> (last visited Oct. 4, 2006) (indicating that this was the ninth proposed amendment to the "Body of the Constitution" put forth by the Virginia convention).

¹⁶¹ 1 ELLIOT'S, *supra* note 123, at 325-27. Among the amendments proposed by the New Hampshire convention was one that would state, "no standing army shall be kept up in time of peace, unless with the consent of three fourths of the members of each branch of Congress; nor shall soldiers, in time of peace, be quartered upon private houses, without consent of the owners." *Id.* at 326.

¹⁶² The New York convention declared:

That the people have a right to keep and bear arms; that a well-regulated militia, including the body of the people capable of bearing arms, is

Rhode Island¹⁶⁴ all specifically addressed the dangers pre-

the proper, natural, and safe defence [sic] of a free state.

That the militia should not be subject to martial law, except in time of war, rebellion, or insurrection.

That standing armies, in time of peace, are dangerous to liberty, and ought not to be kept up, except in cases of necessity; and that at all times the military should be under strict subordination to the civil power.

That, in time of peace, no soldier ought to be quartered in any house without the consent of the owner, and in time of war only by the civil magistrate, in such manner as the laws may direct.

1 ELLIOT'S, *supra* note 123, at 328 (emphasis omitted).

¹⁶³ The North Carolina ratifying convention proposed adoption of a Declaration of Rights, to include the following:

17th. That the people have a right to keep and bear arms; that a well regulated militia composed of the body of the people, trained to arms, is the proper, natural and safe defence [sic] of a free state. That standing armies in time of peace are dangerous to Liberty, and therefore ought to be avoided, as far as the circumstances and protection of the community will admit; and that in all cases, the military should be under strict subordination to, and governed by the civil power.

18th. That no soldier in time of peace ought to be quartered in any house without the consent of the owner, and in time of war in such manner only as the Laws direct.

19th. That any person religiously scrupulous of bearing arms ought to be exempted upon payment of an equivalent to employ another to bear arms in his stead.

BILL OF RIGHTS, *supra* note 67, at 182, 216. That same convention also proposed that the body of the Constitution be amended so "[t]hat no standing army or regular troops shall be raised or kept up in time of peace, without the consent of two thirds of the members present in both houses." 4 ELLIOT'S, *supra* note 123, at 245.

¹⁶⁴ The Rhode Island convention declared, among other things,

That the people have a right to keep and bear arms; that a well-regulated militia, including the body of the people capable of bearing arms, is the proper, natural, and safe defence [sic] of a free state; that the militia shall not be subject to martial law, except in time of war, rebellion, or insurrection; that standing armies, in time of peace, are dangerous to liberty, and ought not to be kept up, except in cases of necessity; and that, at all times, the military should be under strict subordination to the civil power; that, in time of peace, no soldier ought to be quartered in any house without the consent of the owner, and in time of war only by the civil magistrates, in such manner as the law directs.

1 ELLIOT'S, *supra* note 123, at 335. Consistent with this declaration, the Rhode Island ratifying convention recommended that the Constitution be amended to state,

sented by a standing army, even though each voted to ratify the Constitution.

In step with the deliberations taking place in the state legislatures and ratifying conventions, a debate raged in the public-at-large between those who supported adoption of the Constitution (*i.e.*, the Federalists) and those who opposed its ratification (*i.e.*, the Anti-Federalists). The Federalists, who "emphasized order, tradition, the natural distinctions among men, and social harmony,"¹⁶⁵ wrote articles in the country's newspapers, published pamphlets, and gave speeches to rally the public to the new form of government.¹⁶⁶ The most famous of these pro-Constitution efforts were *The Federalist Papers*, a series of letters published by Alexander Hamilton, James Madison, and John Jay in New York area newspapers between October 1787 and May 1788, that were later brought out in book form to influence the delegates to the New York ratifying convention in Poughkeepsie.¹⁶⁷ In recognition of the low place that standing armies occupied with the public, Madison and Hamilton specifically addressed the fear that the Constitution would allow the central government to create a liberty threatening military force.¹⁶⁸

James Madison defended Congress's ability to raise and support troops under the Constitution as a power necessary to

[a]s standing armies, in time of peace, are dangerous to liberty, and ought not to be kept up, except in cases of necessity; and as, at all times, the military should be under strict subordination to the civil power, that, therefore, no standing army or regular troops shall be raised or kept up in time of peace.

Id. at 336.

¹⁶⁵ KOHN, EAGLE, *supra* note 24, at 12.

¹⁶⁶ See, e.g., THE ESSENTIAL FEDERALIST: A NEW READING OF THE FEDERALIST PAPERS (Quentin P. Taylor ed., 1998); THE FEDERALIST PAPERS.

¹⁶⁷ Morris, *supra* note 27, at 55 ("The *Federalist* comprised eighty-five letters, of which fifty-one are attributed to Hamilton, twenty-nine to Madison, and five to John Jay . . .").

¹⁶⁸ See Levin, *supra* note 20, at 1049 ("The popular appeal of [the Anti-Federalists'] arguments [against a national standing army] can be gauged by the considerable effort which the authors of the *Federalist Papers* devoted to responding to this . . . critique.").

the public good.¹⁶⁹ Stating that "[s]ecurity against foreign danger [] is . . . an avowed and essential object of the American Union," Madison claimed that the power to raise armies and equip fleets "is involved in the power of self-defence [sic]" and must be undertaken in times of peace in order to deter, and be prepared for, attacks from other nations.¹⁷⁰ Nevertheless, Madison acknowledged that even a necessary standing force was still a dangerous instrumentality, and thus, "an object of laudable circumspection and precaution."¹⁷¹ Madison claimed that the Constitution had taken this danger into account and guarded against it in three ways. First, because the Constitution would generally create a united America, the country would "exhibit[] a . . . forbidding posture, to foreign ambition," making it far less likely that it would need a large standing force to protect itself than if the nation were to split into independent states or confederacies.¹⁷² Second, Madison drew attention to that part of the Constitution that imposed a limit of two years on the term for which revenue could be appropriated to the support of the army.¹⁷³ And third, Madison trumpeted the fact that the Constitution provided for a well-armed militia officered by men chosen from the respective states that would be able, in his view, to overwhelm any standing army that the national government could ever put in the field.¹⁷⁴

¹⁶⁹ THE FEDERALIST NO. 41 (James Madison).

¹⁷⁰ THE FEDERALIST NO. 41 (James Madison), *supra* note 95, at 224; *see also* James Wilson, Speech at a Public Meeting in Philadelphia (Oct. 6, 1787), PENNSYLVANIA PACKET, reprinted in FRIENDS OF THE CONSTITUTION: WRITINGS OF THE "OTHER" FEDERALIST 1787-1788, at 102, 104 (Colleen A. Sheehan & Gary L. McDowell eds., 1998) ("I do not know a nation in the world, which has not found it necessary and useful to maintain the appearance of strength in a season of the most profound tranquility.").

¹⁷¹ THE FEDERALIST NO. 41 (James Madison), *supra* note 95, at 225.

¹⁷² *Id.* at 226; *see also* THE FEDERALIST NO. 4 (John Jay), *supra* note 95, at 25 ("Wisely therefore do they consider union and a good National Government, as necessary to put and keep them in *such a situation*, as, instead of *inviting* war, will tend to repress and discourage it. That situation consists in the best possible state of defence [sic], and necessarily depends on the Government, the arms, and the resources of the country.") (emphasis added).

¹⁷³ THE FEDERALIST NO. 41 (James Madison).

¹⁷⁴ THE FEDERALIST NO. 46 (James Madison). This last point echoed what

In his contributions to *The Federalist Papers*, Alexander Hamilton mounted a more strident defense of Congress's power to raise and support armies under the new Constitution than had James Madison.¹⁷⁵ As a general proposition, Hamilton asserted that the national government must have the unlimited power to raise armies and build and equip fleets, prescribe rules for the regulation of both, direct each in their operations, and provide for their support.¹⁷⁶ Hamilton defended his position on the basis that "it [wa]s impossible to foresee or define the extent and variety of national exigencies" that may arise or the means that may be necessary to meet those challenges.¹⁷⁷ Therefore, in Hamilton's opinion, "no constitutional shackles can wisely be imposed on the power" of the central government "to preside over the common defence [sic]."¹⁷⁸ Hamilton was adamant that the national government must "be clothed with all the powers requisite to the

Madison had argued in the Federal Convention, see *supra* note 126 and accompanying text, and in the Virginia state ratifying convention, see BILL OF RIGHTS, *supra* note 67, at 194 ("The most effectual way to guard against a standing army . . . is to give the general government full power to call forth the militia, and exert the whole natural strength of the Union, when necessary."); see also A Foreign Spectator [Nicholas Collin], *An Essay on the Means of Promoting Federal Sentiments in the United States*, INDEPENDENT GAZETTER, September 18, 21, 28, 1787, reprinted in FRIENDS, *supra* note 170, at 51 ("The happy situation of America will generally guard her against long and severe wars—but should any such happen; even the power of a veteran army could not subdue a patriotic militia ten times its number, and rendered perfectly military in the course of such war.").

¹⁷⁵ See Levin, *supra* note 20, at 1049 ("Hamilton, who would make support of a sizable national army a central theme of his political career, premised his position on the bellicose view of human nature disputed by the anti-federalists.") (footnote omitted).

¹⁷⁶ THE FEDERALIST NO. 23 (Alexander Hamilton).

¹⁷⁷ THE FEDERALIST NO. 23 (Alexander Hamilton), *supra* note 95, at 126; see also THE FEDERALIST NO. 26 (Alexander Hamilton), *supra* note 170, at 140-41 (arguing that it was necessary to invest in the legislative authority the power to look after the national defense so as not "to embarrass the Government and endanger the public safety").

¹⁷⁸ THE FEDERALIST NO. 23 (Alexander Hamilton), *supra* note 95, at 126; see also THE FEDERALIST NO. 25, (Alexander Hamilton) *supra* note 170, at 135 (claiming that because the potential threats were common dangers, "the means of guarding against [them], ought in like manner, to be the objects of common councils, and of a common treasury").

complete execution of its trust," to include "the formation, direction, [and] support of the national forces."¹⁷⁹

Moving beyond his general defense of a national military establishment, Alexander Hamilton, in recognition of the importance of the issue to the ongoing debate, went to considerable effort to respond to the criticism that the Constitution did not explicitly guard against the existence of standing armies in time of peace. In *The Federalist No. 8*, Hamilton drew a contrast between what the military establishment would look like under the new Constitution and what he predicted it would look like if the states failed to unite under a strong central government.¹⁸⁰ As for the first circumstance, Hamilton argued that any standing army formed under the Constitution would be small in size due to the fact that the European powers that presented the greatest danger of invasion were a significant distance from the United States and did not have forces in their new world colonies that could "give us any dangerous annoyance."¹⁸¹ A national standing army of modest size, in Hamilton's opinion, would present little threat to the civil liberties of the larger and more powerful body of the people.¹⁸² In contrast to that state of affairs, Hamilton

¹⁷⁹ THE FEDERALIST NO. 23 (Alexander Hamilton), *supra* note 95, at 127 (emphasis omitted). In listing what he called "defects" in the existing federal system under the Articles of Confederation, Hamilton said that the ability of the Congress to raise armies was "merely a power of making requisitions upon the States, for quotas of men." THE FEDERALIST NO. 22 (Alexander Hamilton), *supra* note 95, at 119. He claimed that this practice in the course of the Revolutionary War "was found replete with obstructions to a vigorous, and to an economical system of defence [sic]." *Id.*; see also THE FEDERALIST NO. 23 (Alexander Hamilton), *supra* note 95, at 128 ("[W]e must discard the fallacious scheme of quotas and requisitions, as equally impracticable and unjust.").

¹⁸⁰ THE FEDERALIST NO. 8 (Alexander Hamilton).

¹⁸¹ *Id.* at 47. In this regard, Hamilton compared the situation of the United States to that of Great Britain, whose geographic separation and strong marine force enabled it to exist with a relatively small army. *Id.*

¹⁸² *Id.*; see also THE FEDERALIST NO. 28 (Alexander Hamilton), *supra* note 95, at 152 ("[I]n a confederacy, the people, without exaggeration, may be said to be entirely the masters of their own fate. Power being almost always the rival of power; the General Government will, at all times, stand ready to check the usurpations of the State Governments, and these will have the same disposition towards the General Government.").

claimed that the frequent war and constant apprehension that would follow any "wreck of the General Confederacy" would inevitably produce large standing armies in the states that would be placed under the authority of an ever increasingly powerful executive, with the result being "the same engines of despotism which have been the scourge of the old world."¹⁸³

In *The Federalist No. 24*, Alexander Hamilton again explicitly addressed the issue of the Constitution allowing the formation of a national standing army in time of peace.¹⁸⁴ Calling the opposition's objections to the creation of such a force "vehement and pathetic declamation," Hamilton made much of the fact that the Constitution lodged "the whole power of raising armies in the [l]egislature," as opposed to the executive.¹⁸⁵ Characterizing the former as "a popular body, consisting of the representatives of the people periodically elected,"¹⁸⁶ Hamilton, like James Madison, called attention to the fact that the Constitution specifically forbid the appropriation of money for the support of the army for any period longer than two years.¹⁸⁷ Thereafter, somewhat contradicting

¹⁸³ THE FEDERALIST NO. 8 (Alexander Hamilton), *supra* note 95, at 43, 45.

¹⁸⁴ THE FEDERALIST NO. 24 (Alexander Hamilton).

¹⁸⁵ *Id.* at 131 (emphasis omitted). In *The Federalist No. 29*, Hamilton suggested that the opposition's arguments on the issue of a national standing army were "the inflammatory ravings of chagrined incendiaries, or distempered enthusiasts." THE FEDERALIST NO. 29 (Alexander Hamilton), *supra* note 95, at 158.

¹⁸⁶ THE FEDERALIST NO. 24 (Alexander Hamilton), *supra* note 95, at 131; *see also* THE FEDERALIST NO. 26 (Alexander Hamilton), *supra* note 95, at 140, 142 (claiming that conditioning the power to raise and support troops upon the consent of the Congress was "the ultimate point of precaution" and those who argued otherwise possessed "a zeal for liberty more ardent than enlightened"); THE FEDERALIST NO. 28 (Alexander Hamilton), *supra* note 95, at 151 ("[I]t is a full answer to those who require a more peremptory provision against military establishments in time of peace, to say that the whole power of the proposed Government is to be in the hands of the representatives of the people. This is the essential, and, after all, only efficacious security for the rights and privileges of the people, which is attainable in civil society.").

¹⁸⁷ THE FEDERALIST NO. 24 (Alexander Hamilton); *see also* *An American Citizen [Tench Coxe]: An Examination of the Constitution of the United States*, PHILADELPHIA INDEPENDENT GAZETTEER, Sept. 26-29, 1788, reprinted in *FRIENDS*, *supra* note 170, at 473 ("A standing army is not granted or intended, for there can be no provision for its continuing three years, much less for its permanent establishment. . . . It will be under all the restrictions which wisdom and jealousy

the picture he had attempted to paint in *The Federalist* No. 8,¹⁸⁸ Hamilton argued that the danger of invasion from Europe, the threat of British and Spanish colonies in the new world, and "the ravages and depredations of the Indians" on the Western frontier all evidenced "the impropriety of a constitutional interdiction" against standing armies and "the necessity of leaving the matter to the direction and prudence of the [l]egislature."¹⁸⁹

The final issue taken up by Alexander Hamilton in his defense of the military provisions in the Constitution was the function that the militia was to serve under the new system of government. Hamilton had both criticism and praise for the militia. His criticism took the form of an assertion that the militia could never serve as a "natural bulwark" against the dangers represented by European powers and the Native Americans.¹⁹⁰ Claiming that the belief that the militia could be the primary source of defense for the country "had like to have lost us our independence," Hamilton argued that "war . . . [wa]s "a science to be acquired and perfected by diligence," and thus, "[t]he steady operations of war against a regular and disciplined army, can only be successfully conducted by a force of the same kind."¹⁹¹

As for his praise for the institution, Hamilton stated that the militia could serve as an integral part of the country's military establishment as long as it was "well-regulated."¹⁹² Essential components of "a well-regulated militia" were a uniformity in organization and discipline, and this, as Hamilton pointed out, was secured in the new Constitution by its confiding "the regulation of the militia to the direction of the

can suggest, and the original grant of the supplies must be made by the House of representatives, the immediate delegates of the people.").

¹⁸⁸ See *supra* note 181 and accompanying text.

¹⁸⁹ THE FEDERALIST NO. 24 (Alexander Hamilton), *supra* note 95, at 134; see also THE FEDERALIST NO. 28 (Alexander Hamilton), *supra* note 170, at 151 (adding "insurrections and rebellions" to the list of dangers that might necessitate a military response from the national government).

¹⁹⁰ THE FEDERALIST NO. 25 (Alexander Hamilton), *supra* note 95, at 138.

¹⁹¹ *Id.* at 138-39.

¹⁹² See THE FEDERALIST NO. 29 (Alexander Hamilton), *supra* note 95, at 154.

National authority.”¹⁹³ Additionally, Hamilton claimed that the more effective the militia was in meeting the country’s defense needs, the less inducement and pretext there would be for the central government to raise and keep a standing army.¹⁹⁴ Finally, Hamilton cast the militia as offering protection against any threat that a national army might present to civil liberties. He claimed that as long as there was a well-regulated militia with its officers appointed by each of the respective states, its members would stand as a force “ready to defend their own rights and those of their fellow-citizens.”¹⁹⁵

Acting in opposition to the efforts of the Federalists were the Anti-Federalists,¹⁹⁶ who generally objected to what they saw as the “tendency and intention of the Constitution . . . to establish one complete national government, which would de-

¹⁹³ *Id.*; see also George Washington, Circular to the States, June 14, 1783, reprinted in *FRIENDS*, *supra* note 170, at 20 (“The Militia of this Country must be considered as the Palladium of our security, and the first effectual resort in case of hostility; It is essential therefore, that the same system should pervade the whole; that the formation and discipline of the Militia of the Continent should be absolutely uniform, and that the same species of Arms, Accoutrements and Military Apparatus, should be introduced in every part of the United States”) (this letter was published on March 15, 1787, in the Providence *United States Chronicle*).

¹⁹⁴ See *THE FEDERALIST* NO. 29, *supra* note 95, at 154-55.

¹⁹⁵ *Id.* at 157.

¹⁹⁶ The Anti-Federalists usually denied that this label was apt or just, and instead claimed they were the true Federalists. See 1 Storing, *supra* note 83, at 9; see also *BILL OF RIGHTS*, *supra* note 67, at 60-61 (speaking before the First Congress, Elbridge Gerry noted that Anti-Federalists complained that they were the ones in favor of a federal government, whereas their opponents were in favor of a national one.). The Anti-Federalists saw true federalism as meaning that the states were primary, equal to each other, and in possession of the main weight of political power—the very characteristics that they feared the Constitution would undermine. See 1 Storing, *supra* note 83, at 9. However, “the term ‘federal’ had acquired a specific ambiguity that enabled the Federalists not merely to take but to keep the name.” *Id.* Arising out of the pre-1787 debate over the degree to which the general government (*i.e.*, the instrumentality of the federation *per se*) under the Articles of Confederation was to be empowered to act, “one was ‘federal’ or ‘anti-federal’ according to his willingness or unwillingness to strengthen or support the institutions of the federation.” *Id.* This distinction was carried into the debate over the Constitution, so that a Federalist favored strengthening the power of the national government to act, while an Anti-Federalist opposed the same. See *id.* at 9-10.

stroy or undermine the states."¹⁹⁷ This Anti-Federalist position "was localist, republican and committed to the maximum actual participation of the [individual] in the governance of the community."¹⁹⁸ It claimed there was "an inherent connection" between the maintenance of the power and autonomy of the states and "the preservation of individual liberty."¹⁹⁹ As smaller political entities than the proposed national government, the states were better constituted to engender the voluntary attachment of the people, which would, in turn, beget voluntary obedience to the laws.²⁰⁰ Additionally, the states as smaller republics would be able to secure genuine responsibility of the government to the people and form the kind of citizens who would maintain a republican government.²⁰¹

¹⁹⁷ 1 Storing, *supra* note 83, at 10 (citation omitted). This is not to suggest, however, that the Anti-Federalists were a unified block. *See id.* at 5 (noting that although there was "agreement about many points of opposition to the Constitution," "it is not possible to read far among the Anti-Federal writings without being struck by an extraordinary heterogeneity"); Levin, *supra* note 20, at 1043 (explaining that the Anti-Federalists represented "not a single political program or theory, but a whole host of sectional and individual points of view").

¹⁹⁸ Levin, *supra* note 20, at 1044; *see also* David C. Williams, *Civic Republicanism and the Citizen Militia: The Terrifying Second Amendment*, 101 YALE L.J. 551, 576 (1991) ("Anti-Federalist[s] . . . feared that the new central government would be dominated by the Few, distant from the people, and lustful for power.").

¹⁹⁹ 1 Storing, *supra* note 83, at 15; *see also* AKHIL R. AMAR, *THE BILL OF RIGHTS* 11 (1998) (describing Anti-Federalists' fears that aristocrats "in Congress would disdain their lowly constituents, who would in turn lose faith in the national government. . . . [thus driving that government] to rule through corruption, force, and fear—with monopolies and standing armies—rather than through mutual confidence").

²⁰⁰ 1 Storing, *supra* note 83, at 16-17. The Anti-Federalists asserted that "because of the extensive and heterogeneous nature of the society it was to govern, [the federal government created by the new Constitution] would have to act tyrannically." BAILYN ET. AL., *supra* note 56, at 311.

²⁰¹ 1 Storing, *supra* note 83, at 16-17. As for the first point, the Anti-Federalists acknowledged that some form of representative government was necessary, but they believed that there should be as much similarity as possible between the representative body and the people at large in order to maximize effective and thoroughgoing responsibility from the one to the other. *See id.* at 17. As for the character of the citizenship, the Anti-Federalists feared that the peoples of the thirteen American states lacked the shared manners, sentiments, and interests that were necessary to forge a government that could represent the diverse interests of the people, and thus, they feared a constant clashing of opinions and the

As a constituent part of their greater desire for smaller, more accountable and responsive republics, the Anti-Federalists generally held the idea of a national standing army in contempt, while bestowing praise upon the citizen militia.²⁰² As seen earlier, the Anti-Federalists' fear of standing armies was raised many times over in the state ratifying conventions.²⁰³ Beyond this, like the opposing efforts of Madison and Hamilton in *The Federalists Papers*, some Anti-Federalists took their objections to a national standing army directly to the people through the public press.²⁰⁴ For example, in an article published in the *New York Journal* on November 8, 1787, the Anti-Federalist "A Son of Liberty" listed a standing army as one of the "curses" that would "be entailed on the people of America, by [the] preposterous and new-fangled system" of government proposed by the Constitution.²⁰⁵ More pointedly, this author cast a standing army as a "bane to freedom, and support of tyrants, and their pampered minions; by which almost all the nations of Europe and Asia, have been enslaved."²⁰⁶

Other Anti-Federalists penned disparaging remarks about the prospect of a national standing army equal to those offered by "A Son of Liberty." "Cincinnatus," who published a series of

continual striving of one against the other. See *id.* at 19-20.

²⁰² See KOHN, EAGLE, *supra* note 24, at 82-83; 1 Storing, *supra* note 83, at 20.

²⁰³ See *supra* notes 153-64 and accompanying text.

²⁰⁴ The Anti-Federalists had difficulty making themselves heard, in part, because much of the press was closed to them. See BAILYN ET. AL., *supra* note 56, at 313 (noting that "out of a hundred or more newspapers printed in the late [17]80s, only a dozen supported the Anti-Federalists").

²⁰⁵ *Objections by a Son of Liberty*, NEW YORK JOURNAL, Nov. 8, 1787, reprinted in 6 Storing, *supra* note 83, at 34 (emphasis omitted).

²⁰⁶ *Id.* Another curse listed by "A Son of Liberty" was that of the various state militias being dragged all over the country to quell insurrections and "to subdue their fellow citizens, who dare to rise against the despotism of government." *Id.* at 35; see also Letter from The Federal Farmer (Oct. 9, 1787), reprinted in 2 Storing, *supra* note 83, at 233 (reporting that in one of the pamphlets published by the Anti-Federalist, "The Federal Farmer," the author declared that the general government under the Constitution would not be able to "carry all the powers proposed to be lodged in it into effect, without calling to its aid a military force, which must very soon destroy all elective governments in the country, produce anarchy, or establish despotism").

essays during the ratification debates that were widely circulated in the northern states, said in one of his writings that standing armies were something "which all free nations have studiously avoided," were "the source and security of tyranny," and had been "condemn[ed]" by "all great political writers."²⁰⁷ In the first of three essays published by "A New Hampshire Farmer," the author repeated language that had been used many times over when he declared that "[s]tanding armies are dangerous *in time of peace to the liberties of a free people*."²⁰⁸ Remarking that he "ha[d] both wintered and summered with an army," the New Hampshire Farmer noted that "[a]n army, either in peace or war, is like the locust and caterpillars of Egypt; they bear down all before them—and many times . . . have been used as an engine to destroy the liberties of a people, and reduce them to the most abject slavery."²⁰⁹ Writing under the pseudonym "A Columbian Patriot," the Massachusetts Anti-Federalist Mercy Otis Warren proclaimed that "[s]tanding armies have been the nursery of vice and the bane of liberty from the Roman legions . . . to the planting [of] the British cohorts in the capitals of America."²¹⁰ Another citizen of Massachusetts, John DeWitt, challenged the Federalist claim that a standing army was needed in time of peace to maintain an appearance of strength by asking,

[w]here is the standing army in the world, that . . . hath been, in time of peace, brightened and burnished for the sake

²⁰⁷ *Cincinnatus, Essay of November-December 1787*, reprinted in 6 Storing, *supra* note 83, at 16. The essays by Cincinnatus were written as attacks on a speech delivered by the Federalist James Wilson to the citizens of Philadelphia on October 6, 1787. See *id.* at 5. Among other things, Mr. Wilson had claimed that the subject of a standing army was nothing more than "a topic of popular declamation." *Id.* at 16 (quoting JOHN BACH MCCASTER & FREDERICK D. STONE, *PENNSYLVANIA AND THE FEDERAL CONSTITUTION* 149 (1787-1788)).

²⁰⁸ *Essays by a Farmer*, FREEMAN'S ORACLE, NEW HAMPSHIRE ADVERTISER, Jan.-June 1788, reprinted in 4 Storing, *supra* note 83, at 207.

²⁰⁹ *Id.* The Farmer went on to argue that an organized and well-armed militia would be all that was necessary to offer the country "sufficient security." *Id.*

²¹⁰ *Observations on the New Constitution, and on the Federal and State Conventions by A Columbian Patriot*, reprinted in 4 Storing, *supra* note 83, at 270-71, 277.

only of maintaining an appearance of strength, without being put to a different use, without having had a pernicious influence upon the morals, the habits, and the sentiments of society ²¹¹

Further to the south, the Maryland Anti-Federalist "A Farmer" published an essay in the *Maryland Gazette* on March 28, 1788, declaring that "[a]s Alexander dissolved the gordian knot with the sword—so I fear a standing army will simplify the governments of America"²¹² Echoing comments made by George Mason at the Federal Convention,²¹³ "A Farmer" went on to say that only a free people personally connected to the powers of government would be able to defend themselves from outside dangers, as "the people will never fight (if they can help it) for representatives, taxes, and rags."²¹⁴ Next door to Maryland, in one of his many essays that were published in the (Philadelphia) *Independent Gazetteer*, "Philadelphienis" feared that a national standing army would

be made up of profligate idle ruffians, whose prowess will chiefly consist in feats of cruelty exercised on their innocent fellow citizens; but in facing a foreign foe, they will prove themselves to be a body of mean cowards . . . better acquainted with plundering their country than fighting for its protection.²¹⁵

This author also decried the demise of the state militia under the new Constitution and said that it was absurd to think of that body defending the country, as the people under the Con-

²¹¹ *Essays of John Dewitt*, AMERICAN HERALD, Oct.-Dec. 1787, reprinted in 4 Storing, *supra* note 83, at 39. According to DeWitt, the Congress ought to have the power to raise an army only when the country had been invaded by enemies, and he explicitly called for such a reservation to be included in a proposed Bill of Rights. *Id.*

²¹² *Essays by a Farmer*, MARYLAND GAZETTE, Mar. 28, 1788, reprinted in 5 Storing, *supra* note 83, at 45.

²¹³ See *supra* note 134 and accompanying text.

²¹⁴ *Essays by a Farmer*, MARYLAND GAZETTE, Mar. 28, 1788, reprinted in 5 Storing, *supra* note 83, at 46.

²¹⁵ *Essays of Philadelphienis*, INDEPENDENT GAZETTER, Nov. 1787- Apr. 1788, reprinted in 3 Storing, *supra* note 83, at 111.

stitution would be "broken by oppression" and "disaffected," and thus, would never "voluntarily fight for [the] government" ²¹⁶

One of the most important Anti-Federalist writers of the ratification period was "Brutus," who published a series of sixteen essays in the *New York Journal* from October of 1787 through April of 1788 which paralleled and confronted many of the pro-Constitution arguments raised in *The Federalist Papers*.²¹⁷ Along with attacking many other of the Federalists' arguments in favor of the Constitution, Brutus, like the Anti-Federalists quoted above, warned against the danger of a national standing army in time of peace. In his first essay, which set out his broad objections to the new Constitution, Brutus expressed the fear that the national government that was to be created would not have the support of the people, and thus, would have to "establish[] an armed force to execute [its] laws at the point of the bayonet."²¹⁸ Brutus claimed that using a standing army for such a purpose was a sure sign of a despotic government, as those forces "have always proved the destruction of liberty, and [are] abhorrent to the spirit of a free republic."²¹⁹

In the eighth essay of his series, which was published on January 10, 1788, Brutus offered a pointed critique of the Constitution giving Congress the power to borrow money on the credit of the United States and to raise and support armies.²²⁰ Brutus joined these two issues because he saw both powers as being unlimited, and thus, vesting in the general government dangerous control over all the wealth and force of the union.²²¹ In regards to the latter power specifically, Brutus argued that the ability to keep up a national standing army in time of peace ought to be restricted on the grounds that such

²¹⁶ *Id.*

²¹⁷ *Essays of Brutus*, NEW YORK JOURNAL, Oct. 1787-Apr. 1788, reprinted in 2 Storing, *supra* note 83, at 358.

²¹⁸ *Id.* at 371.

²¹⁹ *Id.* at 370 (alteration in original).

²²⁰ *See id.* at 405-08.

²²¹ *See id.* at 405-06.

was "dangerous and improvident."²²² In support of this position, Brutus quoted extensively from a speech delivered by William Pulteney in the British House of Commons, in which Pulteney spoke in favor of a motion to reduce the size of the British army.²²³ In that speech Pulteney said, among other things, that "a standing army of any kind[] . . . is a terrible thing[]" and that it was "indeed impossible that the liberties of the people in any country can be preserved where a numerous standing army is kept up."²²⁴ Recalling the experience of the House of Commons being put under the control of Oliver Cromwell's military, Pulteney said that any large standing army will remain submissive to civilian authority only so long as that authority "does nothing to disoblige their favourite [sic] general; but when that case happens, I am afraid, that in place of the parliament's dismissing the army, the army will dismiss the parliament."²²⁵ Brutus used Pulteney's words to conclude that a standing army was "in the highest degree dangerous to the . . . happiness of the community," and thus, the national "government should [not] be empowered to do that which if done, would tend to destroy public liberty."²²⁶

²²² *Id.* at 406.

²²³ See *id.* at 407-08. "Brutus" said of his use of Mr. Pulteney's speech, that it was "so full to the point, and so much better than any thing I can say, that I shall be excused for inserting it." *Id.* at 407.

²²⁴ *Id.*

²²⁵ *Id.* at 408; see also Britannia.com, Monarchs: Oliver Cromwell (1649-1658 AD), <http://www.britannia.com/history/monarchs/mon48.html> (last visited Oct. 4, 2006).

²²⁶ Brutus, *Essay of January 17, 1778*, NEW YORK JOURNAL, reprinted in 2 Storing, *supra* note 83, at 408. In his ninth essay, "Brutus" wrote that the truth of the proposition that standing armies are "dangerous to the liberties of a people" could be "confirmed by the history of almost every nation in the world." *Id.* at 409. In his tenth essay, published on January 24, 1788, "Brutus" argued that standing armies presented the two specific dangers of being used by duly elected representatives to support themselves "in any usurpations of power" and of the army overthrowing the government and establishing another "according to the pleasure of [its] leader." *Id.* at 413. While "Brutus" acknowledged the need to maintain a small number of troops to garrison important frontier posts and guard arsenals, he did not think that such a limited body of soldiers would be the equivalent of keeping a standing army, and he proposed that their numbers be only that amount that was necessary for the job. *Id.* at 416. He also would have

Despite the efforts of the Anti-Federalists, the Constitution became the law of the land on June 21, 1788, when New Hampshire became the ninth state to ratify the new form of government.²²⁷ Shortly thereafter, Virginia, the most populous state, and New York, the largest, ratified the Constitution as well.²²⁸ Although the Anti-Federalists had failed to block ratification, their efforts were not without effect. Largely because of the objections they raised in the ratification debates, the ink did not even have time to dry on the Constitution before efforts were underway to amend the document.²²⁹

C. Adopting the Bill of Rights

Anti-Federalists' agitation against the Constitution in the state ratifying conventions and the popular press gave rise to the general expectation, if not the promise by its Federalists supporters, that the Constitution would be revisited in the short-term.²³⁰ When the First Congress convened in 1789, James Madison took upon himself the task of compiling, editing and presenting to the Congress the numerous proposed amendments that had come out of the ratification process.²³¹

allowed the raising of troops to ward off a threatened attack or invasion but said that, in times of peace, no troops should be raised "without the assent of two thirds of the members, composing both houses of the legislature." *Id.*

²²⁷ The Avalon Project at Yale Law School, *Ratification of the Constitution by the State of New Hampshire; June 21, 1788*, <http://www.yale.edu/lawweb/avalon/const/ratnh.htm> (last visited Oct. 4, 2006).

²²⁸ See The Avalon Project at Yale Law School, *Ratification of the Constitution by the State of Virginia; June 26, 1788*, <http://www.yale.edu/lawweb/avalon/const/ratva.htm> (last visited Oct. 4, 2006); The Avalon Project at Yale Law School, *Ratification of the Constitution by the State of New York; July 26, 1788*, <http://www.yale.edu/lawweb/avalon/const/ratny.htm> (last visited Oct. 4, 2006); DONALD B. DODD, *HISTORICAL STATISTICS OF THE STATES OF THE UNITED STATES: TWO CENTURIES OF THE CENSUS, 1790-1990*, 93, 467 (1993).

²²⁹ See, e.g., 1 Storing, *supra* note 83, at 3 ("[T]he Constitution that came out of the deliberations of 1787 and 1788 was not the same Constitution that went in; for it was accepted subject to the understanding that it would be amended immediately to provide for a bill of rights.").

²³⁰ See BAILYN ET. AL., *supra* note 56, at 313; LEONARD W. LEVY, *ORIGINS OF THE BILL OF RIGHTS* 30-32 (1999); MALCOLM, *supra* 37, at 159; Uviller & Merkel, *supra* note 75, at 151; see also Cooper, *supra* note 37, at 132-33.

²³¹ See MALCOLM, *supra* note 37, at 159. The ratification conventions of Massa-

In order to alleviate Anti-Federalists' fears regarding the new form of government, Madison had declared himself open to amending the Constitution in ways that were "without danger" in the Virginia ratifying convention debates.²³² Consistent with that caveat, Madison took care to see that the structure and authority of the national government were "as little touched as possible."²³³ The amendments he advanced, for the most part, reflected limitations on the power of the federal government to interfere with civil liberties that were arguably already implicit in the new Constitution.²³⁴ Many Anti-Federalists were incensed by this turn of events.²³⁵

James Madison presented his list of constitutional amendments to the House of Representatives on June 8, 1789.²³⁶ In-

chusetts, South Carolina, New Hampshire, Virginia, New York and North Carolina all endorsed various amendments to the Constitution before the First Congress convened in 1789. AMAR, *supra* note 199, at 14.

²³² ELLIOT'S, *supra* note 123, at 626-27. See LEVY, *supra* note 230, at 32, 38-39; see also AMAR, *supra* note 199, at 14 (commenting on "James Madison's efforts to find common ground with, and to ease the fears of, Anti-Federalist critics of the Philadelphia Constitution").

²³³ Letter from James Madison to Edmund Randolph (June 15, 1789) in 12 THE PAPERS OF JAMES MADISON 219 (Charles F. Hobson & Robert Rutland eds., 1979); see also MALCOLM, *supra* note 37, at 159 (stating that Madison explained to Thomas Jefferson a few days after submitting his amendments that "he had selected those rights for inclusion which were unexceptional and most likely to win approval"); KOHN, EAGLE, *supra* note 24, at 196 ("Madison maneuvered the Bill of Rights through the First Congress precisely to head off many of the more dangerous structural amendments that Antifederalists hoped would reverse the Federalist victory in 1788"). See generally BILL OF RIGHTS, *supra* note 67, at 118 (quoting Letter from Pierce Butler to James Iredell (Aug. 11, 1789)) ("A few *milk-and-water* amendments have been proposed by Mr. M[adison] . . . I suppose it was done to keep his promise with his constituents . . . ; but, if I am not greatly mistaken, he is not hearty in the cause of amendments.") (alteration in original).

²³⁴ See BAILYN, ET. AL., *supra* note 56, at 315; Finkelman, *supra* note 96, at 119-121, 133.

²³⁵ Finkelman, *supra* note 96, at 121, 132; LEVY, *supra* note 230, at 39-40.

²³⁶ MALCOLM, *supra* note 37, at 159.

Madison introduced what would have been twelve, not ten, amendments. One, never ratified, would have forbidden states as well as the federal government to violate "the equal rights of conscience, or the freedom of the press, or the trial by jury in criminal cases." Another was passed by Congress, lay dormant for 200 years, and was ratified in 1992. It is now the Twenty-Seventh Amendment, and it requires that an election must

cluded within Madison's proposals were two that are particularly relevant to the present discussion; *i.e.*, the provisions that would eventually evolve into the Second and Third Amendments to the U.S. Constitution. The item that what would become the Second Amendment proposed that Article I to the Constitution be amended, to wit:

The right of the people to keep and bear arms shall not be infringed; a well armed, and well regulated militia being the best security of a free country: but no person religiously scrupulous of bearing arms, shall be compelled to render military service in person.²³⁷

After being presented to the full House, this amendment was referred to a special committee, which reported back to the House with a slightly different version that inverted the first two clauses of Madison's original proposal.²³⁸ This revised amendment was debated by the House on August 17, with various alterations and additions being proposed.²³⁹ For example, Aedanus Burke, congressman from South Carolina, mounted an unsuccessful effort to require a two-thirds majority in both Houses of Congress to raise a national standing army in time of peace, as such was "dangerous to public liberty."²⁴⁰ After

take place between any Congressional vote to raise its own pay and the actual increase in salary.

Garrett Epps, *The Bill of Rights*, 82 OR. L. REV. 517, 519 (2003) (citing Letter from James Madison to Thomas Jefferson (Oct. 17, 1788), in 11 THE PAPERS OF JAMES MADISON 295, 297-99 (Charles F. Hobson & Robert A. Rutland eds, 1977) (footnote omitted)).

²³⁷ MALCOLM, *supra* note 37, at 159 (quoting 12 THE PAPERS OF JAMES MADISON 220 (Charles F. Hobson & Robert Rutland eds., 1979)). Madison's proposal would have placed the amendment between Clauses Three and Four of Section Nine to the First Article. *Id.*

²³⁸ *Id.* at 160. Madison was a member of the committee. LEVY, *supra* note 230, at 37. The committee proposed inserting the following language between the second and third clauses of Article I, Section 9 to the Constitution: "A well regulated militia, composed of the body of the people, being the best security of a free State, the right of the people to keep and bear arms shall not be infringed, but no person religiously scrupulous shall be compelled to bear arms." MALCOLM, *supra* note 37, at 160 (quoting 1 ANNALS OF CONGRESS 750 (1789-1791)).

²³⁹ See MALCOLM, *supra* note 37, at 160-62.

²⁴⁰ Finkelman, *supra* note 96, at 139-40; MALCOLM, *supra* note 37, at 163.

further debate and after having reached an agreement that all the amendments would be added to the end of the Constitution rather than embedded within it,²⁴¹ the House passed a resolution on August 24 containing its adopted amendments.²⁴² Among these was an amendment that was identified as "Article the Fifth," which stated:

A well regulated militia, composed of the body of the People, being the best security of a free State, the right of the People to keep and bear arms, shall not be infringed, but no one religiously scrupulous of bearing arms, shall be compelled to render military service in person.²⁴³

The House of Representatives' "Article the Fifth" was debated by the Senate on September 4, at which time a motion similar to the one offered by Congressman Burke was made to amend the resolution to, among other things, call for avoiding a standing army in time of peace.²⁴⁴ This motion, like Burke's, was defeated, but the Senate did significantly shorten the House's version of the amendment by dropping the language that would have allowed individuals to avoid military service due to religious scruples.²⁴⁵ On September 9, the Senate

Burke also wanted to add that "in all cases the military shall be subordinate to the civil authority." BILL OF RIGHTS, *supra* note 67, at 172. On the same day that Burke advanced his motion, Elbridge Gerry of Massachusetts asserted that the purpose of the militia was "to prevent the establishment of a standing army, the bane of liberty." *Id.* at 186.

²⁴¹ See LEVY, *supra* note 230, at 37-38; MALCOLM, *supra* note 37, at 161.

²⁴² See, e.g., BILL OF RIGHTS, *supra* note 67, at 85, 173, 209.

²⁴³ *Id.* at 173.

²⁴⁴ *Id.* at 173-74. This motion declared such forces "dangerous to Liberty" and also called for a two-thirds majority in both Houses of Congress to raise a standing army in time of peace, for the military to be strictly subordinate to the "civil Power," and for no soldier to be "inlisted [sic] for any longer term than the continuance of the war." *Id.*

²⁴⁵ *Id.* at 174. This version read: "A well regulated militia, being the best security of a free State, the right of the people to keep and bear arms, shall not be infringed." *Id.* There is no record to explain why the Senate dropped the religious exemption from military service from the amendment. See Finkelman, *supra* note 96, at 139. See generally SPITZER, *supra* note 136, at 27 (referring to the "limited debate in the First Congress surrounding the Second Amendment" and

passed a resolution to amend the Constitution to read, "A well regulated militia, being necessary to the security of a free State, the right of the people to keep and bear arms, shall not be infringed."²⁴⁶ The House accepted the Senate's version on September 21, and the full Congress entered into an agreed resolution to that effect on September 25.²⁴⁷ The Second Amendment, along with nine of the other twelve amendments that were sent to the states, was ratified on December 15, 1791.²⁴⁸

Although the ultimate meaning of the Second Amendment remains in dispute more than two centuries after its passage,²⁴⁹ there is no doubt but that its adoption was a manifestation of the pro-militia, anti-standing army point-of-view that was prevalent at the time.²⁵⁰ The amendment explicitly af-

the "brief debate about and modest wording changes in the amendment's four versions"). In the House, Elbridge Gerry argued that the religious exemption ought to be available only to individuals belonging to a religious sect that objected to its members bearing arms. See BILL OF RIGHTS, *supra* note 67, at 186. Gerry feared that the national government might define a general religious exemption so broadly that it would engineer the demise of the militia. See *id.*

²⁴⁶ BILL OF RIGHTS, *supra* note 67, at 176.

²⁴⁷ *Id.* at 176, 180.

²⁴⁸ THE FEDERALIST PAPERS READER 29 (Frederick Quinn ed., 1993); LEVY, *supra* note 230, at 43. "The first two proposed amendments, which concerned the number of constituents for each Representative and the compensation of Congressmen, were not ratified." *Bill of Rights*, The National Archives Experience, at http://www.archives.gov/national-archives-experience/charters/bill_of_rights.html (last visited Oct. 4, 2006).

²⁴⁹ See, e.g., MALCOLM, *supra* note 37, at 136. Some read the Second Amendment as guaranteeing a right to arms only within the context of a state regulated militia. See, e.g., Finkelman, *supra* note 96, at 140-42; Uviller & Merkel, *supra* note 75, at 148-49; WILLIAMS, *supra* note 100, at 5. Others, who emphasize the Second Clause of the Amendment, claim that it protects a personal right to arms. See, e.g., LEVY, *supra* note 230, at 133-49. Fortunately, it is not necessary to flesh out this debate and choose sides for our present purposes. See e.g., Epps, *supra* note 236, at 521 ("The Second Amendment is so vague that it defies the efforts of the best legal minds to give it a definite meaning."). The drafting and ratification of the Second Amendment is important to this article because of the relationship between that act and the polity's view in late eighteenth century America of the place that the military occupied in our constitutional structure.

²⁵⁰ See AMAR, *supra* note 199, at 56 ("If the [Second A]mendment is not about the critical difference between the vaunted 'well-regulated Militia' of 'the people'

firms that the states are able to arm their militias in the event the national government fails to do so,²⁵¹ and it implicitly reflects a preference for civilian control and decentralized power over the means of employing political violence.²⁵² Writing one hundred years after its adoption, Justice Story put the explicit and implicit in the form of the emphatic when he said that the Second Amendment's protection of the militia "offers a strong moral check against the usurpation and arbitrary power of rulers, and will generally, even if these are successful in the first instance, enable the people to resist and triumph over them."²⁵³

The movement of what would ultimately become the Third Amendment to the U.S. Constitution through the First Congress largely mirrored the route taken by the Second Amend-

and the disfavored standing army, it is about nothing."); Uviller & Merkel, *supra* note 75, at 149 ("[T]he *personal right* expressed in the neoclassical language of the Second Amendment was understood by principal draftsman James Madison to serve the interests of the commonweal[th] by buttressing *community security* and reducing the sway of a dangerous, potentially usurpatory standing army."); WILLIAMS, *supra* note 100, at 40 ("[T]he largely Anti-Federalist voice of the Second Amendment . . . is a voice that worries principally about federal tyranny and finds safety in a popular militia."). In the House debate leading up to its adoption, Elbridge Gerry offered his uncontradicted explanation of the purpose of the Second Amendment to be "to secure the people against the mal-administration of the Government" and "to prevent the establishment of a standing army, the bane of liberty." BILL OF RIGHTS, *supra* note 67, at 186.

²⁵¹ WILLIAMS, *supra* note 100, at 52; Finkelman, *supra* note 96, at 118, 146; AMAR, *supra* note 199, at 50. Delegates to a number of the state ratifying conventions were adamant that the states should retain the ability to arm their militias if the federal government did not. See AMAR, *supra* note 199, at 46-47, 50-51; SPITZER, *supra* note 136, at 24-25. Federalists argued that it was unnecessary to amend the Constitution for this purpose as the states retained a concurrent power under the ratified Constitution to arm the militia if they saw the need. See WILLIAMS, *supra* note 100, at 47, 51.

²⁵² See Millett, *supra* note 135, at 101; WILLIAMS, *supra* note 100, at 64.

²⁵³ 2 JOSEPH STORY, COMMENTARIES ON THE CONSTITUTION OF THE UNITED STATES 646 (5th ed. 1994); see also H. Richard Uviller & William G. Merkel, *Scottish Factors and the Origins of the Second Amendment: Some Reflections on David Thomas Konig's Rediscovery of the Caledonian Background to the American Right to Arms*, 22 LAW & HIST. REV. 169, 170 n.2 (2004) ("Our argument is that the Second Amendment responded to the anti-federalists' classically republican fears of centrally controlled standing armies, and the dangers they posed to balanced, constitutional rule.").

ment. The original version of the former was submitted to the House of Representatives by James Madison on June 8, 1789,²⁵⁴ and reported out of committee, in slightly different form, on July 28.²⁵⁵ With little debate,²⁵⁶ the House passed a resolution on August 24, that would amend the Constitution by adding an "Article the Sixth," which stated: No soldier shall, in time of peace, be quartered in any house without the consent of the owner, nor in time of war, but in a manner to be prescribed by law.²⁵⁷ The provision was quickly adopted as submitted by the Senate on September 9,²⁵⁸ and then both chambers entered into an agreed resolution on the matter on September 25.²⁵⁹ The amendment, along with the remainder of the Bill of Rights, was ratified in December of 1791.²⁶⁰

The Third Amendment, like the Second, arose out of a tradition that sought to "protect[] civilian values against the threat of an overbearing military."²⁶¹ In 1689, the English Bill of Rights, in a single provision, objected to the Crown "raising and keeping a standing army . . . in time of peace without consent of Parlyament [sic] and quartering soldiers contrary to law."²⁶² The Declaration of Independence, in addition to protesting the presence of the King's army in a time of peace,

²⁵⁴ Madison proposed that the following appear in Article 1, Section 9, between the Third and Fourth clauses: "No soldier shall in time of peace be quartered in any house without the consent of the owner; nor at any time, but in a manner warranted by law." BILL OF RIGHTS, *supra* note 67, at 207.

²⁵⁵ *See id.* at 207 ("No soldier shall in time of peace be quartered in any house without the consent of the owner, nor in time of war but in a manner to be prescribed by law.").

²⁵⁶ In keeping with his habit, Elbridge Gerry moved to amend the provision to explicitly designate a civil magistrate as the authority on the law in such an instance as is anticipated by the Third Amendment, but he was defeated by a vote of thirty-five to thirteen. *See id.* at 219.

²⁵⁷ *Id.* at 209.

²⁵⁸ *See id.* at 210.

²⁵⁹ *See id.* at 215.

²⁶⁰ *See supra* note 248 and accompanying text.

²⁶¹ Amar, *supra* note 199, at 59.

²⁶² *See* BILL OF RIGHTS, *supra* note 67, at 217. The English Bill of Rights had been preceded sixty-one years earlier by The Petition of Right of 1628, which had complained that the Crown was billeting soldiers in private homes "against the laws and customs of this realm." CUNLIFFE, *supra* note 69, at 31.

listed the "Quartering [of] large bodies of armed troops among [the colonists]" as one of its "repeated injuries and usurpations."²⁶³ Contemporaneous with ratifying the Constitution, the state conventions of New Hampshire, New York, North Carolina, Rhode Island and Virginia all called for amendments that would both identify standing armies as dangerous to liberty and prohibit the quartering of troops in civilian homes without consent of the owners.²⁶⁴ The deep anxiety about a potentially abusive military reflected in these events was a prime motivation behind adoption of the Second and Third Amendments to the Constitution.²⁶⁵ The military amendments sought to forestall, respectively, civilian society being dominated by a standing army in peacetime "either openly or by subtle insinuation."²⁶⁶

III. THE MILITARY ESTABLISHMENT THROUGH THE 1790S

The years from 1790 to 1800 were spent trying to fashion the new government called for by the Constitution and the Bill of Rights. Along with disputes regarding the financial program of the central government,²⁶⁷ the size and scope of the national military establishment was a constant source of debate in the political life of 1790s America.²⁶⁸ Consistent with the tenor of the debate from the 1760s through the 1780s, any argument over the national military establishment in the last decade of the eighteenth century eventually raised the issue of the danger of standing armies in time of peace. Prime examples of efforts to strike a balance between the security needs of the nation and the antipathy that many felt toward regular military forces can be found in the country's response to the Indian wars in the early 1790s, the effort to reform the militia in

²⁶³ THE DECLARATION OF INDEPENDENCE, *supra* note 69, at para. 2, 16. See *supra* note 70 and accompanying text.

²⁶⁴ See BILL OF RIGHTS, *supra* note 67, at 181-83, 215-16.

²⁶⁵ AMAR, *supra* note 199, at 46-47.

²⁶⁶ *Id.* at 59.

²⁶⁷ See KOHN, EAGLE, *supra* note 24, at 117, 131, 166, 196, 259; WEIGLEY, *supra* note 144, at 17.

²⁶⁸ KOHN, EAGLE, *supra* note 24, at 296.

1792, the Whiskey Rebellion of 1794, and the Quasi War with France at the end of the eighteenth century.

A. *The Indian Wars*

As settlers in the 1780s and early 1790s moved into the western sections of Pennsylvania and Virginia, and into the Kentucky District, conflicts between those settlers and the Indian tribes of the Northwest Territory began to occur with increasing frequency.²⁶⁹ This, in turn, created significant political pressure on the national government to do something about the deteriorating security situation on the frontier.²⁷⁰ Efforts by the federal government to subdue the northwest Indian tribes dominated the debate over the place of the military in American life from 1790 until the defeat of the latter at the Battle of Fallen Timbers in August of 1794.²⁷¹

In anticipation of action on the frontier against the Indian tribes, President Washington asked Congress in 1790 to authorize an additional army regiment of 800 infantrymen.²⁷² The House assented in short order, but the Senate resisted furiously and approved the creation of only a new battalion of 400 extra men.²⁷³ Many senators objected to the administration's request because they suspected a plan to establish a standing army, they feared the expense of creating and maintaining such a force, and they believed that the militia could meet the security needs on the frontier.²⁷⁴ Consistent with the latter sentiment, a military campaign led by General Josiah Harmer marched into the Ohio territory in the Fall of 1790 with 300 regular army troops and 1200 militiamen.²⁷⁵ After its defeat by the tribes in October, reports came flowing in which blamed the failure of the campaign on the poor fighting quality of the

²⁶⁹ See *id.* at 92-95. For a map of the relevant area, see *id.* at 105.

²⁷⁰ See *id.* at 92-95.

²⁷¹ *Id.* at 156-57.

²⁷² *Id.* at 100.

²⁷³ *Id.* at 100-01.

²⁷⁴ *Id.* at 101.

²⁷⁵ *Id.* at 103-05.

militia.²⁷⁶ Those forces were said to be composed largely of poorly equipped substitutes, made up of old men and young boys who were "hardly able to bear arms."²⁷⁷ Moreover, according to reports of regular army officers on the scene, when fired upon by the Native Americans for the first time, the militia men threw down their arms and ran from the field.²⁷⁸ Based upon its performance in the Harmar campaign, "[n]ever again would a Federalist administration willingly use militia against Indians, indeed for any military crisis."²⁷⁹

Following on the heels of General Harmar's defeat, the Washington Administration returned to Congress in January 1791 seeking authorization and appropriations for a 3000-man military force to carry out an invasion of the Northwest Territory, north from Cincinnati, Ohio.²⁸⁰ The force would be composed of a regiment of volunteer cavalry made up of westerners from the frontier, a new regiment of regular soldiers, and a regiment of "levies" or volunteers enlisted by the government for the duration of the campaign.²⁸¹ The levies were "to be raised, officered, and controlled by the national government under regular army discipline."²⁸² As federal volunteers on short enlistments, the levies provided "an institutional alternative to the politically unpopular standing army and the militarily unreliable state militia,"²⁸³ and promised to be far

²⁷⁶ *Id.* at 106-08.

²⁷⁷ *Court of Inquiry on General Harmar*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 20, 20-21 (Walter Lowrie & Walter S. Franklin eds., 1832) (testimony of Major Ferguson); *see also id.* at 23 (testimony of Ensign Morgan) (The militia "were very disorderly, and very inattentive to their duty, and some appearances of mutiny among them, with both officers and men"); KOHN, EAGLE, *supra* note 24, at 107-08 (The militia in the Harmar campaign was "[c]omposed of substitutes, poorly armed, badly equipped, and officered by inexperienced men . . . , [and] uncontrollable, disobedient, and at crucial times openly mutinous").

²⁷⁸ *See Court of Inquiry on General Harmar*, *supra* note 277, at 21 (testimony of Major Ferguson); *id.* at 23 (testimony of Lieutenant Hartshorn).

²⁷⁹ KOHN, EAGLE, *supra* note 24, at 108.

²⁸⁰ *See id.* at 109-10.

²⁸¹ *See id.*

²⁸² *Id.* at 110.

²⁸³ *Id.* at 137.

cheaper than regular forces.²⁸⁴ After learning of the killing of twelve men, a woman, and two children by Native Americans in the area of Fort Harmar, Congress acceded to the administration's requests for its new force, but the results did not improve.²⁸⁵ The ensuing military campaign, led by Major General Arthur St. Clair, met with defeat in November of 1791, suffering 600 dead and over 300 wounded.²⁸⁶

Staggered by its military losses but undeterred, the Washington Administration proposed a new campaign into the Northwest Territory in January of 1792.²⁸⁷ The administration asked Congress to authorize and support a new force of 5000 men divided among four regiments of regular infantry, one of riflemen and a battalion each of artillery and cavalry.²⁸⁸ Many legislators associated with the emerging Republican Party took the opportunity "to accuse the Federalists of plotting the introduction of a standing army and introducing policies to strengthen [Secretary of Treasury Alexander] Hamilton's financial edifice with added federal debts."²⁸⁹ The effort initially failed to gain the necessary support by a single vote in the Senate.²⁹⁰ Senator Benjamin Hawkins of North Carolina, "who well symbolized the administration's critics," expressed "respect" for Secretary of War Henry Knox, but stated that the Secretary "appears . . . to be anxiously desirous of

²⁸⁴ *Id.* at 110.

²⁸⁵ *Id.*

²⁸⁶ See *id.* at 116. At the time of the decisive attack by the Native Americans upon his forces, General St. Clair had approximately fourteen hundred men at his disposal. *Causes of the Failure of the Expedition Against the Indians, in 1791, Under the Command of Major General St. Clair*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 36, 37 (Walter Lowrie & Walter S. Franklin eds., 1832). The levies had performed badly due to the fact that their short enlistment period had prevented adequate training. See *id.*; KOHN, EAGLE, *supra* note 24, at 119.

²⁸⁷ See KOHN, EAGLE, *supra* note 24, at 120.

²⁸⁸ *Id.*; WEIGLEY, *supra* note 144, at 18.

²⁸⁹ KOHN, EAGLE, *supra* note 24, at 121. Many western Republicans who hated standing armies nevertheless supported the effort as they could not oppose protection for their home areas. See *id.*

²⁹⁰ *Id.* at 122.

having considerable standing military force.”²⁹¹ Once the administration replaced General St. Clair with a new commanding general and committed to a renewed effort to interest the tribes in a negotiated settlement, Congress authorized the administration to form five 960-man regiments for three-year enlistments.²⁹² This force came to be referred to as the “[L]egion,” based upon the organization of its regiments into integrated fighting units.²⁹³

The Washington Administration spent the remainder of 1792 and all of 1793 rebuilding and reorganizing the regular army and negotiating with the northwest Indian tribes.²⁹⁴ In this lull, during December 1792 and January 1793, the House of Representatives engaged in a heated debate about the nation’s military establishment after Congressman John Steele moved to reduce the army to two regiments.²⁹⁵ In defending his motion, which ultimately failed, Congressman Steele accused the administration of militarism and of violating the principle of rejecting standing armies in peacetime.²⁹⁶ In the midst of the debate and for the first time, James Madison spoke openly against the use of regulars, although he abstained from the final vote.²⁹⁷ At this same time the rising Republican Party wholeheartedly embraced the military issue, as it so squarely represented struggles between national versus state power and standing armies versus “the people” as militia.²⁹⁸

²⁹¹ *Id.* at 123.

²⁹² *Id.* As a concession to the opposition, a provision was adopted that required the discharge of the new regiments whenever peace should be established along the frontier. *Id.*

²⁹³ *Id.* at 126; CUNLIFFE, *supra* note 69, at 50. Each regiment was to have its own cavalry, artillery and infantry. See *Organization of the Army in 1792*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 40, 41 (Walter Lowrie & Walter S. Franklin eds., 1832). This made each “a self-contained force able to act independently with mobility, flexibility, and hitting power.” KOHN, EAGLE, *supra* note 24, at 124.

²⁹⁴ See KOHN, EAGLE, *supra* note 24, at 141-43.

²⁹⁵ See *id.* at 147.

²⁹⁶ *Id.*

²⁹⁷ See *id.* at 148.

²⁹⁸ See *id.* In many states former Anti-Federalists formed the nucleus of the Republican Party and that successor “opposed a national military establishment in

After it was shown beyond doubt in late 1793 that the British were actively encouraging the Indian tribes to wage war against the United States, the public mood changed from one of hoping for a negotiated settlement to one of anticipating war.²⁹⁹ In preparation for the coming hostilities, General Anthony Wayne was ordered to use 2000 frontiersmen as auxiliaries to fill out the Legion army that Congress had approved in early 1792.³⁰⁰ After the Washington Administration learned in June of 1794 about the existence of a new British fort on the Maumee River on American soil, General Wayne was ordered into action.³⁰¹ Wayne moved north into Indian territory in late July, and by October had scored a series of victories over the Indians that “finally broke the Indian hold on the old Northwest and destroyed British influence over the tribes.”³⁰² The major victory came in the Battle of Fallen Timbers on August 20 near the newly erected British fort, when Wayne’s forces routed over 1000 Native Americans and a company of Canadian militia.³⁰³ The hostilities were formally brought to a close with the signing of the Treaty of Greenville on August 3, 1795.³⁰⁴

The Federalists military program of the first half of the 1790s, which revolved largely around the effort to subdue the Indian tribes in the Northwest Territory, was “sold to a suspicious public . . . one element at a time, always as a temporary

the same terms as had their predecessors in the 1780s.” *Id.* at 197.

²⁹⁹ See *id.* at 155-56. Two events in particular contributed greatly to this change in public outlook. First, in October of 1793 three American peace emissaries publicly blamed the British for the failure of their efforts to negotiate a settlement with the Native American tribes. *Id.* at 156. Second, Joseph Brant, leader of the Six Nations, publicly accused British agent Alexander McKee in December of 1793 of encouraging the western tribes to choose war. *Id.* at 155-56.

³⁰⁰ *Id.* at 156. Notwithstanding congressional authorization to raise a 4800 man force, General Wayne’s army had never exceeded 3300 regulars. *Id.* at 176.

³⁰¹ *Id.* at 156. The Maumee River is located in the northwest corner of Ohio. See *id.* at 66.

³⁰² *Id.* at 157.

³⁰³ *Id.*

³⁰⁴ See BAILYN ET. AL., *supra* note 56, at 321-22; The Avalon Project at Yale Law School, *Treaty of Greenville 1795*, <http://www.yale.edu/lawweb/avalon/greenvil.htm> (last visited Oct. 4, 2006).

defensive expedient.”³⁰⁵ Even after public opinion had turned to war in 1794 due to British provocation and Native American depredations along the frontier, James Madison continued to give voice to the complaint that the Federalists were “turning every contingency into a resource for accumulating force in the government.”³⁰⁶ John Adams, Vice President in the Washington Administration and himself an ardent Federalist, saw great danger lurking in the build-up of military forces in preparation for General Wayne’s campaign. In a letter written on May 11, 1794, to Thomas Jefferson, Vice President Adams predicted that “[a]nother War would . . . raise [u]p a many headed and many bellied Monster of an Army to tyrannize over us, totally dissadjust [sic] our present Government, and accelerate the Advent of Monarchy and Aristocracy by at least fifty Years.”³⁰⁷ As exemplified by Madison’s and Adams’ comments and the debates in Congress each time the administration sought an expansion in the regular army, notwithstanding the country’s precarious military situation in the early 1790s, the fear of a standing army drove many to insist upon a showing of absolute necessity for regular forces before those forces would be approved.³⁰⁸ And even then, the permission was generally

³⁰⁵ KOHN, EAGLE, *supra* note 24, at 174-75. The Federalists’ military program also included “coast and harbor forts, a corps of engineers and artillerists to man them, naval vessels, and new arsenals and armories to manufacture and store arms and equipment.” *Id.* at 174. These improvements were “pushed through Congress” as measures to protect the nation against Britain in 1794. *Id.*

³⁰⁶ *Id.* at 175 (quoting a letter written by Madison to Thomas Jefferson in March of 1794).

³⁰⁷ *Id.* at 173 (citing THE ADAMS-JEFFERSON LETTERS: THE COMPLETE CORRESPONDENCE BETWEEN THOMAS JEFFERSON AND ABIGAIL AND JOHN ADAMS 255 (Lester J. Cappon ed., 1959)).

³⁰⁸ See *supra* notes 279, 285, 292, 299-301, and accompanying text. As another example of the tenor of the debates that took place in Congress on the issue of military forces, consider the March 12, 1794, motion by Theodore Sedgwick, congressman from Massachusetts, proposing to raise fifteen regiments of “auxiliary or provisional troops.” KOHN, EAGLE, *supra* note 24, at 220 (quoting 3 ANNALS OF CONGRESS 485 (1793-1795)). Congressman Sedgwick’s plan would arm and equip the men enlisted for these regiments, but would require them to serve on active duty for only twenty-four days a year unless war broke out. *Id.* Congressman Sedgwick argued that the force presented all the advantages of a regular army but without any attendant threat to liberty or significant drain on resources. *Id.*

given with great reluctance and with an expectation that as soon as the necessity passed the forces would be disbanded.³⁰⁹

B. *The Uniform Militia Act of 1792*

On August 10, 1789, President Washington reported to the U.S. Senate that the national military establishment totaled 672 officers and enlisted men.³¹⁰ In the face of the threats presented by the British from the north and the Native Americans on the frontier, the Washington Administration believed that such a small number of troops was not sufficient to meet the country's security needs.³¹¹ However, it was well understood that because of the existing pervasive hostility to a national standing army, Congress was not likely to authorize a larger military force in a time of peace.³¹² Therefore, the Washington Administration settled on reform of the militia as the next best alternative to the creation of a substantial professional army.³¹³ This effort got underway in January of 1790 when Secretary of War Henry Knox presented a plan for the *Organization of the Militia* to the Senate and House of Representatives.³¹⁴

The Republican opposition saw great dangers in the proposed force, with Senator James Monroe claiming that the plan represented "the commencement of a military establishment." *Id.*

³⁰⁹ See WEIGLEY, *supra* note 144, at 18-19; see also *Military Force in 1794*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 68, 68 (Walter Lowrie & Walter S. Franklin eds., 1832).

The force immediately under Major General Wayne is much lessened by the expiration of the services of his troops, and is inferior to the demand of existing circumstances. Unless, therefore, he be [sic] reinforced early in the ensuing spring, the advantages which he has gained in the course of the present year, which ought to be permanently secured, must be in danger of being relinquished.

Id.

³¹⁰ *Military Force in 1789*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 5, 6 (Walter Lowrie & Walter S. Franklin eds., 1832).

³¹¹ See Uviller & Merkel, *supra* note 75, at 151-52.

³¹² See *id.* at 152.

³¹³ See *id.*; CUNLIFFE, *supra* note 69, at 182-83; WEIGLEY, *supra* note 144, at 19.

³¹⁴ *Organization of the Militia*, in CLASS V AMERICAN STATE PAPERS: 1 MILI-

In deference to the political sentiments of the time, Secretary Knox introduced his reform measure by stating that a standing army "cannot in peace be considered as friendly to the rights of human nature" and that "[a]n energetic national militia is to be regarded as the *capital security* of a free republic."³¹⁵ In an effort to secure the latter energetic force, Secretary Knox proposed classing the militia into three groups; *i.e.*, the advanced corps, the main corps, and the reserved corps.³¹⁶ The advanced corps would be made up of all men aged eighteen to twenty and would train from ten to thirty days each year.³¹⁷ The main corps would consist of men aged twenty-one to forty-five and would train four days a year.³¹⁸ The reserved corps would be men aged forty-six to sixty and would assemble twice annually for the inspection of arms.³¹⁹ Wanting to avoid the typical approach of "idle [sic] and worthless" substitutes being employed,³²⁰ Secretary Knox believed that classing was necessary to transform the militia into an effective fighting force.³²¹ His plan was met with "shock and disbelief."³²²

Notwithstanding the fact that the Washington Administration put forth the Knox proposal as an alternative to the creation of a standing army, many saw the evils of that dreaded

TARY AFFAIRS 6, 6 (Walter Lowrie & Walter S. Franklin eds., 1832).

³¹⁵ *Id.* at 7.

³¹⁶ *Id.* at 8. Secretary Knox's plan would have thoroughly nationalized the militia, with the federal government dictating the structure and organization of the body, and taking responsibility for its administration, supply, and training. KOHN, EAGLE, *supra* note 24, at 130.

³¹⁷ *Organization of the Militia*, *supra* note 314, at 8-9.

³¹⁸ *Id.* at 8-10.

³¹⁹ *Id.* Secretary Knox had advanced a very similar plan in his 1786 "Plan for the General Arrangement of the Militia of the United States." CUNLIFFE, *supra* note 69, at 182. George Washington had argued in favor of dividing the militia into classes in 1783. See *Sentiments on a Peace Establishment, 1783*, *supra* note 144, at 24-25.

³²⁰ *Organization of the Militia*, *supra* note 314, at 7.

³²¹ KOHN, EAGLE, *supra* note 24, at 129; see also Uviller & Merkel, *supra* note 75, at 152 (explaining that the Washington Administration believed that an effective militia "required more training and discipline than could possibly be instilled by the states mustering their entire adult male population for a single day or, at best, a few days each year").

³²² KOHN, EAGLE, *supra* note 24, at 130.

institution lurking in the background. The idea of classing the militia was not far removed from the formation of a select militia, and the latter had historically been seen as "a form of standing army liable to be skewed politically and dangerous to liberty."³²³ Additionally, many feared that the annual training that would be required of the advanced corps would "militarize the nation."³²⁴ For example, a Pennsylvania congressman worried that the country's youth would learn only "[i]dleness & extravagance" in military camps, while the author, "A Mechanic," argued in the Philadelphia *Freeman's Journal* that the advanced corps would not be the place for young men who were establishing and fixing their morals to acquire "virtuous principles."³²⁵ Moreover, as was frequently the case with any proposed expansion in the military establishment, the expense that would be incurred "shocked most congressmen and senators."³²⁶ Nevertheless, a congressional committee was appointed in the House to consider Secretary Knox's proposal and that committee began drafting legislation in April of 1790.³²⁷

The House committee quickly abandoned Secretary Knox's class system and, instead, proposed putting eighteen to twenty-four year-olds into special light infantry or rifle companies attached to regular battalions and regiments.³²⁸ These com-

³²³ MALCOLM, *supra* note 37, at 148; see also Uviller & Merkel, *supra* note 75, at 153 (asserting that selection of only the youngest men for extensive training and service "entailed formation of fighting bands less firmly rooted in their communities by family and property than the historic common militia").

³²⁴ KOHN, EAGLE, *supra* note 24, at 131.

³²⁵ *Id.* (quoting Letter from William Irvine to John Nicholson (Feb. 21, 1790) (on file with Historical Society of Pennsylvania); Letter from A Mechanic, in PHILADELPHIA FREEMAN'S JOURNAL, Mar. 17, 1790). The related issue of the status of conscientious objectors also raised a great deal of concern in areas with significant Quaker populations. See *id.* While Secretary Knox was willing to relax the principle of universal obligation for officials, ministers, and possibly some religious sects, he was adamant "that measures of national importance never should be frustrated for the accommodation of individuals." *Id.* (quoting 1 ANNALS OF CONGRESS 2099 (1789-1791)).

³²⁶ *Id.* The expense for the advanced corps alone was estimated to be \$400,000 per year. *Id.*

³²⁷ *Id.*

³²⁸ *Id.*

panies would muster on their own four times a year and with their parent unit in the army twice a year.³²⁹ Unlike the Knox plan, the committee's proposal "made liberal provision for exceptions from service, but most [of the] groups excused (Quakers included) would pay a yearly tax of two dollars."³³⁰ The bill was introduced into the House at the end of its first session in July 1790.³³¹

The House began debating the revised Knox proposal in December of 1790, with every provision being subjected to intense review.³³² "Differences of opinion [quickly arose] on everything from the age of service to the number and purpose of muster days, to the number, rank, and pay of inspectors."³³³ Ultimately, "agreement was impossible on any provision which increased the burden on individual citizens or provided for a specific national standard," and thus, the bill was gradually "stripped of its strongest provisions in order to satisfy the chorus of conflicting views."³³⁴ A new committee was appointed to prepare another version of the bill, but this effort solved nothing and the First Congress disbanded without further action.³³⁵

In January of 1792, following St. Clair's defeat, the Washington Administration proposed raising the 5000 man "Legion" for a new frontier campaign.³³⁶ This action squarely raised the issue of a tradeoff between a regular standing army versus reform of the militia, thus prodding Congress to return to the issue it had abandoned in 1791.³³⁷ This time the debate

³²⁹ *Id.* at 131-32.

³³⁰ *Id.* at 132.

³³¹ *Id.*

³³² *See id.* at 132-33.

³³³ *Id.* at 133. *See id.* ("The Quakers, who flooded Congress with petitions in favor of exemption on religious grounds without payment of a penalty tax, were upset because the new bill did not leave exemptions up to individual states.").

³³⁴ *Id.*

³³⁵ *See id.*

³³⁶ *See supra* notes 292-93, and accompanying text.

³³⁷ *See* KOHN, EAGLE, *supra* note 24, at 120, 134. Commenting on the choice being put before the Congress, Republican James Monroe said, "Anything is preferable to nothing, as it takes away one of the arguments for a standing army." *Id.* at 134 (quoting Letter from James Monroe to Archibald Stuart (Mar. 14,

hinged on the extent of national jurisdiction over the militia, with the end result being a weaker bill than had even been debated in the First Congress.³³⁸ Abandoning any provision for classing, "[t]he Uniform Militia Act of 1792 put all 'free able-bodied white male citizen[s]' from the ages of [eighteen] through [forty-five] in the militia."³³⁹ A lengthy list of exemptions was adopted and the states were given freedom to add to that list.³⁴⁰ The act offered no guarantees for uniformity in training, structure, or equipment,³⁴¹ and because there was no procedure for enforcing the few national guidelines included in the act, "[i]f a state defied the national system, the federal government was impotent to intercede."³⁴²

C. The Whiskey Rebellion

In 1791, the federal government instituted a whiskey excise tax as an integral part of Secretary of Treasury Alexander Hamilton's financial program.³⁴³ Westerners detested this revenue measure for a couple of different reasons. First, the excise was payable only in cash, and cash was something that frontier farmers rarely had in their possession.³⁴⁴ Second, as it was very difficult to transport any product over the mountains that separated the western frontier from the east coast, farmers shipped their grain to the eastern markets in the form of distilled spirits rather than transporting the raw grain itself.³⁴⁵

1792), in *The Papers of James Monroe*).

³³⁸ See *Id.* at 134; CUNLIFFE, *supra* note 69, at 182-83; Uviller & Merkel, *supra* note 75, at 153; WEIGLEY, *supra* note 144, at 20.

³³⁹ KOHN, EAGLE, *supra* note 24, at 135 (alteration in original). See Uniform Militia Act of 1792, ch. 33, 1 Stat. 271, *repealed by* Dick Act, ch. 196, 32 Stat. 775 (1905).

³⁴⁰ Uviller & Merkel, *supra* note 75, at 153; KOHN, EAGLE, *supra* note 24, at 135.

³⁴¹ Uviller & Merkel, *supra* note 75, at 153; KOHN, EAGLE, *supra* note 24, at 135.

³⁴² KOHN, EAGLE, *supra* note 24, at 135; see also WEIGLEY, *supra* note 144, at 21-22.

³⁴³ See BAILYN ET. AL., *supra* note 56, at 320-21; KOHN, EAGLE, *supra* note 24, at 158-61.

³⁴⁴ See Friendship Hill National Historic Site, *The Whiskey Rebellion*, <http://www.nps.gov/archive/frhi/whiskreb.htm> (last visited Oct. 4, 2006).

³⁴⁵ See KOHN, EAGLE, *supra* note 24, at 158-61; Friendship Hill, *The Whiskey*

These grievances against the whiskey tax fed a pre-existing widespread dissatisfaction that many westerners felt towards the federal government³⁴⁶ and sparked an uprising that eventually required President Washington to lead 15,000 troops into western Pennsylvania to put down a rebellion.³⁴⁷

The opening salvo in the Whiskey Rebellion occurred in the summer of 1792 when a mob attacked an excise inspection station in Washington County, Pennsylvania, and threatened to murder the owner for allowing his house to be used for registering stills.³⁴⁸ In the face of pressure from Secretary Hamilton to respond to the mob action with force, and to include military force if necessary, President Washington instead chose to issue a proclamation that called upon all "to refrain and desist from . . . unlawful combinations and proceedings whatsoever having for object or tending to obstruct the operation of the laws."³⁴⁹ The President was worried about inciting events if he responded to the provocation more strongly. He specifically feared the reaction that might follow if he used the regular army against the rebellious faction, telling Secretary Hamilton that if he did so, "there would be a cry at once, 'The cat is let out; We now see for what purpose an Army was raised.'"³⁵⁰

President Washington's proclamation of 1792 did not bring

Rebellion, supra note 344.

³⁴⁶ "Their grievances involved resolving the Indian problems and opening the Mississippi River to navigation." Friendship Hill, *The Whiskey Rebellion, supra* note 344.

³⁴⁷ See KOHN, EAGLE, *supra* note 24, at 139, 169-70. Some sources put the size of the army led by President Washington at a lesser number of troops. See Uviller & Merkel, *supra* note 75, at 156 (13,000 troops); Friendship Hill, *The Whiskey Rebellion, supra* note 344 (12,950 militia men).

³⁴⁸ KOHN, EAGLE, *supra* note 24, at 160. There had been earlier riots against the tax in North Carolina. See *id.*

³⁴⁹ The Avalon Project at Yale Law School, *George Washington: Proclamation of September 15, 1792*, <http://www.yale.edu/lawweb/avalon/president/proclamations/gwproc08.htm> (last visited Oct. 4, 2006). President Washington repeated the proclamation in a joint speech to Congress on November 6, 1792. George Washington, U.S. President (Nov. 6, 1792), in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 18, 19 (Walter Lowrie & Walter S. Franklin eds., 1832).

³⁵⁰ KOHN, EAGLE, *supra* note 24, at 160 (internal quotation marks omitted) (citation omitted).

the dissent to an end. On July 16 and 17 of 1794, a mob attacked the home of John Neville, who served as the excise inspector responsible for collecting the tax, and took David Lenox, a federal marshal for the District of Pennsylvania, prisoner.³⁵¹ Members of Neville's household were injured in the attack on his home and he was forced to flee the county in fear for his life.³⁵² Lenox was detained until he agreed to refuse to execute some of his official duties on behalf of the United States courts.³⁵³ President Washington reacted to these attacks by issuing a proclamation on August 7 which ordered the rebels to disperse by September 1 and "stat[ed] the government's intention to ready the militia."³⁵⁴ At the same time, Secretary of War Knox wrote the governors of four states to alert their troops and the Washington Administration formed a three person federal commission to attempt to negotiate with the insurgents.³⁵⁵ The commission reported at the end of August that enforcement of the excise law was impossible and recommended immediate military mobilization.³⁵⁶

After further unsuccessful attempts on the part of the federal commission to secure assurances that the rebels would obey the law and express loyalty to the United States, Presi-

³⁵¹ *Id.* at 158; The Avalon Project at Yale Law School, *The Whiskey Rebellion*, <http://www.yale.edu/lawweb/avalon/president/proclamations/gwproc03.htm> (last visited Oct. 4, 2006) (quoting Proclamation from President George Washington (Aug. 7, 1794), CLAYPOOLE'S DAILY ADVERTISER, Aug. 11, 1794).

³⁵² KOHN, EAGLE, *supra* note 24, at 158; *The Whiskey Rebellion*, Avalon, *supra* note 351.

³⁵³ *See The Whiskey Rebellion*, Avalon, *supra* note 351.

³⁵⁴ KOHN, EAGLE, *supra* note 24, at 164 (citation omitted); *The Whiskey Rebellion*, Avalon, *supra* note 351. President Washington recounted the series of events that led to his August 7th proclamation in a speech to Congress on November 19, 1794. George Washington, U.S. President (Nov. 19, 1794), in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 24, 24-25 (Walter Lowrie & Walter S. Franklin eds., 1832).

³⁵⁵ *See* KOHN, EAGLE, *supra* note 24, at 164-65 (the commission was made up of U.S. Senator James Ross, U.S. Attorney General William Bradford, and Pennsylvania Supreme Court Justice Jasper Yeates).

³⁵⁶ *Id.* at 166. President Washington decided to proceed with force in an August 24 meeting with Hamilton and Secretary of State Edmund Randolph, but overt actions toward that end were delayed in order to allow negotiations to proceed. *See id.* at 167.

dent Washington issued a proclamation justifying the federal government's "resort to military coercion" at the end of September. 35³⁵⁷ At the beginning of October, the President personally led militia from New Jersey, Maryland, and Virginia into the rebellious part of Pennsylvania.³⁵⁸ "The march of the militia [was] almost anticlimactic."³⁵⁹ The army encountered a civilian population that was "terrified by the approach of a crusading [force] and a void left by the escape of most of the rebel leaders."³⁶⁰ Some arrests were made, but only a few insurgents were taken to Philadelphia for trial.³⁶¹ "Of these, only two were convicted of treason and both received pardons from the President."³⁶²

As a postscript to the rebellion, although the rebels faded away in the face of the show of force made by the federal government, the performance of the militia in the march was held up to universal scorn.³⁶³ Two-thirds of the men mobilized did not possess their own guns and many were totally ignorant of firearms.³⁶⁴ The troops that marched on western Pennsylvania were, on the whole, an undisciplined force which "looted, drank heavily, and beat civilians randomly."³⁶⁵ The Washington Administration tried to use this dismal performance as a justification for reform of the militia.³⁶⁶ Secretary Knox, in a

³⁵⁷ See *id.* at 167-69 (quoting THE WRITINGS OF GEORGE WASHINGTON FROM THE ORIGINAL MANUSCRIPT SOURCES, 1745-1799 (John Clement Fitzpatrick et al. eds., (1931-1944); Proclamation from President George Washington (Sept. 25, 1794), available at <http://wyllie.lib.virginia.edu:8086/perl/toccer-new?id=WasFi33.xml&images=images/modeng&data=/texts/english/modeng/parsed&tag=public&part=390&division=div1> (last visited Oct. 4, 2006)).

³⁵⁸ KOHN, EAGLE, *supra* note 24, at 169-70.

³⁵⁹ *Id.* at 170. The only deaths to occur as a result of the march "came from a pistol going off accidentally and a drunken brawl that ended with a fatal bayonet wound." Michael A. Bellesiles, *The Second Amendment in Action*, in THE SECOND AMENDMENT IN LAW AND HISTORY 48, 57 (Carl T. Bogus ed., 2000) (citing THOMAS P. SLAUGHTER, THE WHISKEY REBELLION: FRONTIER EPILOGUE TO THE AMERICAN REVOLUTION 192-206 (1986)).

³⁶⁰ KOHN, EAGLE, *supra* note 24, at 170.

³⁶¹ *Id.*

³⁶² *Id.* (citation omitted).

³⁶³ *Id.* at 135; Bellesiles, *supra* note 359, at 57.

³⁶⁴ Bellesiles, *supra* note 359, at 57; KOHN, EAGLE, *supra* note 24, at 135.

³⁶⁵ Bellesiles, *supra* note 359, at 57.

³⁶⁶ George Washington, U.S. President (Nov. 19, 1794), in CLASS 1 AMERICAN

report submitted to Congress on December 10, 1794, lamented the absence of any state enforcement of requirements for militiamen to arm themselves and he suggested that a select militia "taken from the classes least injurious to the industry of the community" should replace the general militia.³⁶⁷ Although Secretary Knox's proposal had some support, too many believed that more federal control over the militia threatened the constitutional powers of the states and that the states could be persuaded to reform their militias without federal intervention, and thus, the effort went nowhere.³⁶⁸

D. The Quasi-war with France

Following the adoption in 1795 of both Jay's Treaty³⁶⁹ and the Treaty of Greenville,³⁷⁰ pressure mounted for reductions to be made in the country's military forces.³⁷¹ In 1796, the Fourth Congress engaged in "a particularly heated and partisan" debate over the structure and size of the nation's military establishment.³⁷² The Republicans were able to reduce the number of frigates from six to three, but had little success attacking programs for arsenals, stores, or the rest of the navy.³⁷³ The biggest battle, however, developed over the size of the regular army.

On March 25, 1796, a House of Representatives committee formed to study alterations in the army recommended that the

STATE PAPERS: 1 FOREIGN RELATIONS 24, 25 (Walter Lowrie & Walter S. Franklin eds., 1832). President Washington called on Congress to apply its "full energy" to the task of properly "organizing, arming, and disciplining, the militia." *Id.*

³⁶⁷ *Militia*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 69, 69-71 (Walter Lowrie & Walter S. Franklin eds., 1832).

³⁶⁸ WEIGLEY, *supra* note 144, at 21. See generally CUNLIFFE, *supra* note 69, at 185 ("The basic difficulty . . . was that a casually administered, voluntary militia was militarily useless, while a well-trained, obligatory system was politically unattainable.").

³⁶⁹ See BAILYN ET. AL., *supra* note 56, at 326; *supra* note 90 and accompanying text.

³⁷⁰ See *supra* note 304 and accompanying text.

³⁷¹ KOHN, EAGLE, *supra* note 24, at 182-83.

³⁷² See *id.* at 183.

³⁷³ See *id.*

Legion's authorized strength be reduced to 2000 men.³⁷⁴ Although many Federalists argued that military necessity mandated a larger force, the majority did not agree. In the face of warnings about standing armies and the claim that "Government would be better without any army, as it was always better for Government to rest upon the affections of the people than to be supported by terror," the House accepted the committee's report.³⁷⁵ The Senate agreed to the reduction in force, and thus, "[f]or the first time since 1789, Republicans succeeded in paring down the army's authorized strength, and, by mandating regiments with small companies, reorganizing it into a truly peacetime institution."³⁷⁶ Pressure would soon mount, however, for the country to once again build up its military establishment.

In response to what it perceived to be pro-British policies by the United States government as expressed in Jay's Treaty, the French initiated a series of attacks on American shipping and refused to treat with the U.S. until the new American connection with Britain was broken.³⁷⁷ Following his election in 1796 to the office of President, John Adams called a special session of Congress in May of 1797 to prepare the public for possible hostilities with France and to convince the French

³⁷⁴ *Organization of the Army*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 112, 112 (Walter Lowrie & Walter S. Franklin eds., 1832); KOHN, EAGLE, *supra* note 24, at 184.

³⁷⁵ KOHN, EAGLE, *supra* note 24, at 184-85 (referring to a statement made by Congressman (Republican) William Branch Giles).

³⁷⁶ *Id.* at 185. The most important contributing factors to this outcome were "the absence of military danger and the large budgetary demands of the army." *Id.* An earlier effort in Congress in late 1794 and early 1795 to reduce the authorized strength of the Legion had failed. *See id.* at 176-77. In the midst of that earlier debate James Madison "stimulated some heat by moving to limit the Legion to frontier operations, claiming that it might otherwise become the standing army to execute the laws that Federalists were obviously plotting." *Id.* at 177-78.

³⁷⁷ BAILYN ET. AL., *supra* note 56, at 326-27; *see also France*, in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 748, 748-49 (Walter Lowrie & Walter S. Franklin eds., 1832) (noting that in response to an inquiry from the Senate, Secretary of State Timothy Pickering reported on various complaints received by his department from American citizens regarding, among other things, French ships of war and privateers seizing American vessels and their cargo).

government of American resolve.³⁷⁸ When President Adams pressed for military preparations and negotiations with the French, the Republicans feared the steps were a prelude to war and fought against naval and military measures in both the House and Senate.³⁷⁹ For example, Senator Henry Tazewell of Virginia complained that every action that prepared for war proved to France that the U.S. government was insincere in its effort to negotiate.³⁸⁰ However, in the wake of public disclosure of the "X, Y, Z Affair" in early 1798, in which the French government asked American envoys for a bribe before it would open negotiations to resolve differences, the Congress was forced to act.³⁸¹ As rumors of internal conspiracies against public order and the government abounded, legislation passed and was signed by the President in July that authorized the creation of twelve new regiments of regular troops.³⁸² This result did not come easy.

The first efforts to strengthen the country's defenses in early 1798 met with little opposition.³⁸³ Within a few weeks of the X, Y, Z Affair coming to light, "new laws to prohibit arms exports, expand the coast defense system, and purchase cannon and other arms or establish foundries to manufacture them rolled through Congress with hardly any debate."³⁸⁴ More ambitious measures, however, did not receive such a warm welcome. On April 9, Secretary of War James McHenry asked Congress to authorize three new regiments for the regular army and a Provisional Army of 20,000 men.³⁸⁵ While the Senate

³⁷⁸ John Adams, U.S. President (May 16, 1797), in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 40, 40-41 (Walter Lowrie & Walter S. Franklin eds., 1832); KOHN, EAGLE, *supra* note 24, at 208-09.

³⁷⁹ KOHN, EAGLE, *supra* note 24, at 208-09.

³⁸⁰ *Id.* at 223.

³⁸¹ BAILYN ET. AL., *supra* note 56, at 327; KOHN, EAGLE, *supra* note 24, at 210-11. President Adams, with the assent of Congress, had dispatched Elbridge Gerry, John Marshall, and Charles Cotesworth Pinckney to Paris to negotiate with the French in the summer of 1797. See KOHN, EAGLE, *supra* note 24, at 209.

³⁸² KOHN, EAGLE, *supra* note 24, at 212-14.

³⁸³ *Id.* at 224.

³⁸⁴ *Id.*

³⁸⁵ *Fortifications, Munitions, and Increase of the Army*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 120, 120-21 (Walter Lowrie & Walter S.

accepted the proposal, "[t]he House exploded in anger and vituperation."³⁸⁶ House Republicans saw the effort as one to militarize the country; they questioned whether France truly presented a danger of invasion and argued that "the militia [would] provide[] far better protection than the 'refuse of society'" that would enlist in McHenry's army.³⁸⁷ Many Republicans feared that the Provisional Army would be manned overwhelmingly by zealous Federalists who would then be used for internal control instead of external defense.³⁸⁸ Congress eventually authorized the Provisional Army at the end of May, but in greatly reduced form.³⁸⁹ Instead of the requested 20,000 men, Congress authorized only 10,000 and stipulated that the forces could be raised only if war was declared, the country was invaded, or there was an "imminent danger" of invasion.³⁹⁰

Following a successful effort in June to secure congressional approval to strengthen the volunteer corps and to allow the President to appoint and commission officers to the Provisional Army,³⁹¹ the Federalists returned to Congress in July seeking to increase that force to 50,000 men, with complete mobilization of 12,500.³⁹² However, faced with resistance from moderates in their own party who feared being branded as warmon-

Franklin eds., 1832); KOHN, EAGLE, *supra* note 24, at 224.

³⁸⁶ KOHN, EAGLE, *supra* note 24, at 224.

³⁸⁷ *Id.* at 224-25 (citation omitted).

³⁸⁸ BAILYN ET. AL., *supra* note 56, at 330; KOHN, EAGLE, *supra* note 24, at 225.

In fact, Alexander Hamilton stated that many

Federalists wished to revive 'the spirit of 1776,' to overawe the seditious multitudes infesting the country, 'to keep in check these people,' to 'put arms into the hands of all our friends,' and to cement to the government the 'great body of feeble, deluded, or calculating Neuters who in all Societies join those at last whom they think strongest.

KOHN, EAGLE, *supra* note 24, at 226 (citation omitted).

³⁸⁹ Act of May 28, 1798, ch. 47, 1 Stat. 558.

³⁹⁰ *Id.*

³⁹¹ Act of June 22, 1798, ch. 57, 1 Stat. 569.

³⁹² KOHN, EAGLE, *supra* note 24, at 227. The Federalists also sought "appointment of all [army] officers, . . . Presidential authority to activate . . . non-commissioned officers [for the provisional force], . . . and ten frigates or ships of the line." *Id.* at 227-28.

gers, the Federalists were forced to accept a compromise.³⁹³ The Provisional Army was left at its authorized level of only 10,000 men with no immediate activations, but twelve new regiments of regular troops were authorized to be enlisted for the duration of the "existing differences between the United States and the French Republic, unless sooner discharged."³⁹⁴ This force, which would become known as the "New Army," would entail over 12,000 men and would be raised as soon as circumstances allowed without the need for an invasion or presidential approval.³⁹⁵ From this point forward, the New Army became the focus of Federalists efforts over the previously authorized Provisional Army.³⁹⁶ Their dreams of a large standing army, however, would not come true.

Primarily as a result of the efforts of Secretary McHenry and the recently appointed Inspector General to the Army, Alexander Hamilton, a great deal of time and energy were expended at the end of 1798 and into 1799 in the creation of the New Army, to include the production of arms, promulgation of recruiting instructions, the appointment of officers, and reorganization of ranks and duties.³⁹⁷ The momentum for the force, however, began to bleed away in early 1799 as the threat from France seemed to recede.³⁹⁸ In February, "President [Ad-

³⁹³ *Id.* at 228.

³⁹⁴ Act of July 16, 1798, ch. 76, 1 Stat. 604. The Act also authorized six troops of light dragoons. *Id.*

³⁹⁵ *Id.*; KOHN, EAGLE, *supra* note 24, at 229.

³⁹⁶ KOHN, EAGLE, *supra* note 24, at 239. Secretary McHenry did not give up on the Provisional Army entirely. He proposed in his 1798 annual report to the President that this force be reauthorized at the level of 20,000 men. *Reorganization of the Army*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 124, 126 (Walter Lowrie & Walter S. Franklin eds., 1832).

³⁹⁷ KOHN, EAGLE, *supra* note 24, at 236-37, 240-47. Alexander Hamilton was second in command of the New Army, subordinate to George Washington who was appointed head of the Army. BAILYN ET. AL., *supra* note 56, at 330. See KOHN, EAGLE, *supra* note 24, at 231-33.

³⁹⁸ KOHN, EAGLE, *supra* note 24, at 249; see also CUNLIFFE, *supra* note 69, at 49 ("A letter from a Connecticut citizen printed in the *New London Bee* in May 1799 declared that young farmers would never volunteer 'under Prussian military discipline, or devote their valor to promote the views of ambition or to oppose their country and prosperity with a standing army.'). In March of 1799, the President was authorized by Congress to raise twenty-eight additional regiments

ams] nominated William Vans Murray to reopen negotiations with France" and for the rest of the year, "military ardor steadily diminished."³⁹⁹ Because of his political rivalry with Hamilton and because he was never convinced of the need for the New Army and saw it as a financially and politically costly undertaking, President Adams offered little support for the creation of this body of troops.⁴⁰⁰

Public backing for the New Army was further undermined by a string of reports of soldiers committing criminal acts against civilians.⁴⁰¹ In June of 1799, "marines and cartmen fought a pitched battle on the Albany docks over a horse stabbed by a sergeant."⁴⁰² In August two drunken privates knifed a tavern owner who refused to allow them to gamble in his establishment—"convincing proof," trumpeted the [Philadelphia] *Aurora*, 'that the *privates* who exchange the price of honest industry, for the *humiliating duties* of soldiers, are "villians more desperate, and cut-throats equally bloody minded, with

in the event war ensued or the country was in danger of invasion. See KOHN, EAGLE, *supra* note 24, at 229, 252. This force was referred to as the Eventual Army, but, like the Provisional Army of 1798, it was never organized. See *id.*

³⁹⁹ KOHN, EAGLE, *supra* note 24, at 249. In a compromise with Congress, President Adams nominated Oliver Ellsworth and William R. Davie to join Murray in Europe, but "delay[ed] their departure until France gave explicit assurances that the three would be honorably received." See *id.* at 257, 259. President Adams ordered Ellsworth and Davie to sail for France in October. See *id.* at 259; see also John Adams, U.S. President (Dec. 3, 1799), in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 50, 51 (Walter Lowrie & Walter S. Franklin eds., 1832)) (President Adam's review of his decision to send the ministers to France).

⁴⁰⁰ See KOHN, EAGLE, *supra* note 24, at 237, 243, 247; see also John Adams, U.S. President (Dec. 8, 1798), in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 47, 47-48 (Walter Lowrie & Walter S. Franklin eds., 1832) (stating that the country did "not fear war [with France] in the necessary protection of our rights and honor," President Adams' only specific admonition to Congress was for it to increase the size of the navy).

⁴⁰¹ John Adams, in 1815, said of the New Army:

[It] was . . . unpopular, as if it had been a ferocious wild beast let loose upon the nation to devour it. In newspapers, in pamphlets, and in common conversation [the soldiers] were called cannibals. A thousand anecdotes, true or false, of their licentiousness, were propagated and believed.

KOHN, EAGLE, *supra* note 24, at 256 (second alteration in original) (citation omitted).

⁴⁰² *Id.* at 260 (citation omitted).

the soldiery of the British monarchy."⁴⁰³ The soldiers' antics became such an intense political issue that "[d]uring the 1799 gubernatorial election in Pennsylvania, Republicans published a temporary newspaper called the 'Cannibal's Progress' specifically to list army excesses."⁴⁰⁴

By December 1799, when the Sixth Congress convened, "even confirmed Hamiltonians admitted that the army had become the object of hatred and ridicule."⁴⁰⁵ Virginia representative John Nicholas, a Republican, initiated a legislative attack on the New Army in Congress on January 1, 1800, by moving to repeal the laws authorizing the twelve regiments and the dragoons.⁴⁰⁶ Publicly, the Republicans emphasized the cost of maintaining the troops, the receding French threat, and the ability of the militia to provide the country with an effective defense.⁴⁰⁷ Privately, the Republicans, and some moderate Federalists, feared that some high Federalists desired to use the army for political intimidation or worse.⁴⁰⁸ The New Army would eventually disband in May of 1800, having never enlisted even half of its authorized number.⁴⁰⁹ The eulo-

⁴⁰³ *Id.* (citation omitted). "The next month near Bristol, Pennsylvania, a local law enforcement officer was 'kicked and maltreated' by soldiers when he attempted to arrest a thief who had escaped by enlisting in the army." *Id.* (citation omitted).

⁴⁰⁴ *Id.* (citation omitted); see also CUNLIFFE, *supra* note 69, at 51 ("As late as 1799 an executive head, Secretary of the Treasury Oliver Wolcott, could tell another Federalist that peacetime soldiers, being condemned to inactivity, tended to become morally corrupt.") (citation omitted).

⁴⁰⁵ KOHN, EAGLE, *supra* note 24, at 260.

⁴⁰⁶ *Id.*

⁴⁰⁷ *Id.* at 260-61.

⁴⁰⁸ *Id.* at 262.

⁴⁰⁹ See *Suspension of the Recruiting Service*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 132, 132 (Walter Lowrie & Walter S. Franklin eds., 1832); see also John Adams, U.S. President (Nov. 22, 1800), in CLASS 1 AMERICAN STATE PAPERS: 1 FOREIGN RELATIONS 53, 53 (Walter Lowrie & Walter S. Franklin eds., 1832) (explaining that President Adams reported to Congress that "the officers and soldiers of the temporary army have been discharged"). American and French representatives signed an agreement in September of 1800 that brought the quasi-war to a close. BILYN ET. AL., *supra* note 56, at 330; The Avalon Project at Yale Law School, *France: Convention of 1800*, <http://www.yale.edu/lawweb/avalon/diplomacy/france/fr1800m.htm> (last visited on Oct. 4, 2006).

gy for that force was provided by the presidential election of 1800, with Thomas Jefferson and the Republican Party repeatedly charging that the Federalist had aimed to establish a classic standing army that would threaten civil liberty and impose tyranny.⁴¹⁰

IV. FIRST PRINCIPLES

*There was to be a tiny standing army. . . . There was to be a militia in which every male American was supposed to serve but none w[as] to be given any rigorous training. In time of need an army would be improvised in various ways: by augmenting the regulars, by embodying militia and by raising volunteer regiments.*⁴¹¹

Thus has one historian described what he called "a set of vaguely defined rules" governing the American military establishment at the end of the eighteenth century.⁴¹² As is apparent from the history canvassed in this Article, the reason the country took such an inefficient approach to meeting its security needs was because the polity held it as a testament of faith that a national standing army would be a threat to civil liberties and freedoms.⁴¹³ Americans of this period "viewed a

⁴¹⁰ See KOHN, EAGLE, *supra* note 24, at 268 ("The military became a motif for ridicule and character assassination; Hamilton became Alexander I and on one occasion the *Aurora* parodied Adams, McHenry, and Pickering as officers of the ship '*Administration*' sailing on an 'ocean . . . covered with regimentals.'") (citation omitted). The Republican Party returned to the issue in the 1804 presidential election. See CUNLIFFE, *supra* note 69, at 284. An anonymous pamphleteer charged that the Federalist Party had intended to turn the New Army into a standing army with Alexander Hamilton at its head. *Id.*

⁴¹¹ CUNLIFFE, *supra* note 69, at 52 (emphasis added).

⁴¹² *Id.* at 52.

⁴¹³ KOHN, EAGLE, *supra* note 24, at 282; see also *id.* at 278 (noting that Harrison Gray Otis, U.S. Representative from Massachusetts, said during the demobilization debate of 1800 that "[t]he alarm relative to standing armies has been at least rung a thousand times a year since the first British army was landed in this country") (citing 6 ANNALS OF CONGRESS 305); Baron Friedrich Von Stueben, Letter on the Subject of an Established Militia, and Military Arrangements, Addressed to the Inhabitants of the United States (1784), in ROBERT K. WRIGHT, JR. & MORRIS J. MACGREGOR, JR., SOLDIER-STATESMAN OF THE CONSTITUTION 201,

standing army as the one uncontrollable institution in society."⁴¹⁴ They feared its instinct towards "undermin[ing] civilian authority" and despised its effect of "debauch[ing]" the public virtue.⁴¹⁵ After expelling the redcoats, it would have seemed like a nullification of the revolution for America to erect a new standing military force, even if it was its own.⁴¹⁶

The depth of the public's disdain for a standing army can be seen in a three-part cycle that was emblematic of the political battles of the 1790s over the size and make-up of the national army. The first stage in the cycle was the congressional authorization stage, wherein a highly skeptical Congress would refrain from approving new troops until it perceived the country to be faced with an immediate and concrete external threat.⁴¹⁷ The second, or recruitment, stage would see the executive branch struggle, without great success, to convince enough of the able bodied men of the day to join the army in an

201-06 (1987) (recognizing that a classic standing army was highly improbable in late eighteenth century America, the Baron recommended that the country maintain a small, highly disciplined corps of long-service professionals "composed of your brothers and your sons").

⁴¹⁴ KOHN, EAGLE, *supra* note 24, at 282; *see also* Letter from Thomas Jefferson to David Humphreys (Mar. 18, 1789), in 14 THE PAPERS OF THOMAS JEFFERSON 676, 678 (Julian P. Boyd ed., 1958).

There are instruments so dangerous to the rights of the nation and which place them so totally at the mercy of their governors that those governors, whether legislative or executive, should be restrained from keeping such instruments on foot but in well defined cases. Such an instrument is a standing army.

Id.

⁴¹⁵ KOHN, EAGLE, *supra* note 24, at 282. *See generally* CUNLIFFE, *supra* note 69, at 57 (Dr. Benjamin Rush, former surgeon general of the Continental Army, argued in 1793 that the War Department should have "signs . . . placed over its various doors which read 'an office for butchering the human species,' 'a widow-and orphan-making office,' and so on. .").

⁴¹⁶ WEIGLEY, *supra* note 144, at 17. While it is true that in 1800 there existed a chain of forts along the northern and western frontier manned by an army of 2500 officers and men, such a small constabulary, exiled in frontier garrisons and carefully limited in size and function, was a very different force from the dreaded standing army, in that it did not pose any significant internal political threat. *See* KOHN, EAGLE, *supra* note 24, at 286.

⁴¹⁷ *See supra* notes 289, 303, 385-86, and accompanying text.

effort to raise the troops authorized by Congress.⁴¹⁸ The third and final stage was the decommissioning stage, in which congressionally mandated demobilization of the troops was quick to follow the demise of the threat that initially justified an increase in forces.⁴¹⁹

Congress' management of the country's armed forces in the 1790s satisfied two fundamental responsibilities that had been assigned to it in the Constitution. First, by being intimately involved in decisions about, for example, how large the army was to be and how it was to be organized, the Congress exercised civilian control over the military.⁴²⁰ Second, by requiring the executive to thoroughly justify its requests for troops and to report on its use of the same, the Congress, consistent with the separation of powers principle, provided a check and balance against any misuse of military forces.⁴²¹

The enshrining in the Constitution of the principles of civilian control over the military and separation of powers were the direct result of the trepidation felt by the polity towards a national standing army. When concerns were raised at the

⁴¹⁸ See *supra* notes 284-90, 304, 413, and accompanying text.

⁴¹⁹ See *supra* notes 378-80, 410, and accompanying text.

⁴²⁰ It is the Congress, not the military establishment, that has the power to "declare War," "raise and support Armies," "provide and maintain a Navy," "make Rules for the Government and Regulation of the land and naval Forces," "provide for calling forth the Militia," and "to provide for organizing, arming, and disciplining[] the Militia." U.S. CONST. art. I, § 8, cls. 11, 12, 13, 14, 15, 16; see also MALCOLM, *supra* note 37, at 151-52 ("To ensure that the[] arm[y] remained subject to civilian control, delegates [to the Constitutional Convention] proposed that 'no appropriation of Money to that Use shall be for a longer Term than two Years.'").

⁴²¹ The Congress has the responsibility to decide the size and structure of the military establishment and to decide when the country will go to war, while the President is given the responsibility as commander in chief of using the country's military forces to protect national interests; see also CUNLIFFE, *supra* note 69, at 43 (noting that the Constitution's "military aspects, what might be called the Anglo-American tradition was enshrined in the firm establishment of civil control and in the division of authority between legislature and executive"); MALCOLM, *supra* note 37, at 151 (noting that the delegates to the Constitutional Convention "took great care to create checks and balances that could prevent misuse of military power"). Compare U.S. CONST. art. I, § 8, cls. 11, 12, 13, 14, with U.S. CONST. art. II, § 2, cl. 1.

Constitutional Convention about the danger of the federal government dominating the military establishment, assurances were offered which relied upon the fact that Congress, the people's representatives, would be in control of those forces.⁴²² When the debate moved to the ratification process, Alexander Hamilton defended the Constitution by noting that control over the military was largely vested in "the representatives of the people periodically elected," whose discretion was circumscribed by the Constitution's prohibition against "the appropriation of money for the support of the army for any longer period than two years."⁴²³ Thereafter, the Second and Third Amendments to the Constitution, which arose out of a tradition that sought to "protect[] civilian values against the threat of an overbearing military,"⁴²⁴ reflected, respectively, a preference for civilian control over the means of employing political violence⁴²⁵ and a prohibition against the potentially abusive practice of foisting the burden of maintaining the military on select individuals.⁴²⁶

A third constitutional principle governing our nation's military establishment that has its roots in our founding experience is the need to strike an appropriate balance between military efficiency and the rights of individual service members. Admittedly, the founding generation's concern for individual soldiers was more of an implicit rather than explicit principle in the initial efforts to create and manage the country's military establishment. In the last two decades of the eighteenth century America was, by necessity, struggling to get a structure in place that would enable the general government to look to the collective security needs of the nation, whether to repel an external force or subdue an internal uprising. The vast majority of the polity's attention was, therefore, on the army as an institution, not on the army as a collection of individuals.

⁴²² See 2 RECORDS, *supra* note 86, at 330, 386.

⁴²³ THE FEDERALIST NO. 24 (Alexander Hamilton), *supra* note 95, at 131.

⁴²⁴ AMAR, *supra* note 199, at 59.

⁴²⁵ See *supra* note 252 and accompanying text.

⁴²⁶ See *supra* notes 262-66 and accompanying text.

Although the structure of the military institution dominated the debate at the country's founding, the well-being of individual soldiers was an issue that some felt compelled to raise. During the ratification period following the adjournment of the Federal Convention in Philadelphia, many expressed concern that the military provisions of the new Constitution did not account for the rights of conscientious objectors.⁴²⁷ For example, a minority of the delegates to the Maryland state ratifying convention publicly attacked the Constitution on the grounds that under its terms a person "conscientiously scrupulous of bearing arms" could "be compelled personally to serve as a soldier."⁴²⁸ Furthermore, many also attacked the Constitution on the grounds that it would subject individual militia men to undue hardships.⁴²⁹ The Anti-Federalist letter writer who

⁴²⁷ See BILL OF RIGHTS, *supra* note 67, at 182-83 (stating that the state ratifying conventions of North Carolina, Rhode Island, and Virginia all proposed that individuals religiously scrupulous of bearing arms should be exempt from having to do so upon payment of an equivalent to employ another in their stead); Letter from the Centinel to the People of Pennsylvania (Nov. 5, 1787) (INDEPENDENT GAZETTEER, FREEMAN'S JOURNAL), *reprinted in* 2 Storing, *supra* note 83, at 159 (explaining that in one of his many letters published in Philadelphia newspapers from October 1787 to April 1788, the "Centinel" lamented the fact that the Constitution did not exempt men "of conscientious scruples of bearing arms" from having to serve in a militia that would now be largely controlled by the federal government); Letter from An Old Whig to Mr. Printer, INDEPENDENT GAZETTEER, *reprinted in* 3 Storing, *supra* note 83, at 36 (explaining that an essay written by "An Old Whig" warned that the ability of conscientious objectors to avoid military service would depend "wholly upon the wisdom and discretion of the future legislature of the continent in the framing their militia laws"); *The Address and Reasons of Dissent of the Minority of the Convention of Pennsylvania to Their Constituents*, PENNSYLVANIA PACKET AND DAILY ADVISOR, Dec. 18, 1787, *reprinted in* 3 Storing, *supra* note 83, at 164 (explaining that a minority of the delegates to the state ratifying convention in Pennsylvania derided the Constitution for not containing an exemption from military service for "those persons who are conscientiously scrupulous of bearing arms"); Address of the Albany Antifederal Committee (Apr. 10, 1798), *in* 6 Storing, *supra* note 83, at 123 (stating that the Albany (NY) Antifederal Committee on April 26, 1788, included among its published list of objections to the Constitution the critique that "[m]en conscientiously scrupulous of bearing arms, [would be] made liable to perform military duty").

⁴²⁸ Address of a Minority of the Maryland Ratifying Convention, MARYLAND GAZETTE, May 6, 1788, *reprinted in* 5 Storing, *supra* note 83, at 97.

⁴²⁹ See *id.* at 98 (explaining that the Maryland minority delegates believed it was imperative to restrain the ability of Congress to march the general militia

adopted the name "Centinel" warned that the Constitution gave the federal government the power to "drag[]" men "from [their] families and homes to any part of the continent, and for any length of time."⁴³⁰ Similarly, minority delegates to the Pennsylvania state convention bemoaned the fact that the Constitution gave Congress the power to subject men in the militia "to corporal punishments of the most disgraceful and humiliating kind, and to death itself."⁴³¹ Finally, "Brutus," the Anti-Federalist essay writer in New York, took the "harsh punishment" complaint and applied it to the standing army that he feared the Constitution would allow the federal government to form when he wrote about officers and soldiers being subjected to "punishment so severe that neither . . . dare dispute the orders of his supreme commander."⁴³²

Notwithstanding that there were voices expressing concern for the well-being of individuals in the military, this issue was not a central part of the debate surrounding the creation of the

"beyond the limits of a neighboring state"); *Objections by a Son of Liberty*, NEW YORK JOURNAL, Nov. 8, 1787, reprinted in 6 Storing, *supra* note 83, at 35 ("The militia of New-Hampshire [sic], or Massachusetts, [would be] dragged to Georgia or South-Carolina [sic] to assist in quelling an insurrection of Negroes in those states . . ."); see also BILL OF RIGHTS, *supra* note 67, at 181-83 (explaining that the state ratifying conventions of North Carolina, New York, Rhode Island, and Virginia all called for the militia to be immune from martial law unless and until they were called up by the federal government in time of war, invasion or rebellion).

⁴³⁰ Letter from the Centinel to the People of Pennsylvania (Nov. 5, 1787) (INDEPENDENT GAZETTEER, FREEMAN'S JOURNAL), reprinted in 2 Storing, *supra* note 83, at 159.

⁴³¹ *The Address and Reasons of Dissent of the Minority of the Convention of Pennsylvania to Their Constituents*, PENNSYLVANIA PACKET AND DAILY ADVISOR, Dec. 18, 1787, reprinted in 3 Storing, *supra* note 83, at 164.

⁴³² *Brutus*, Essay of January 10, 1788, NEW YORK JOURNAL, reprinted in 2 Storing, *supra* note 83, at 407; see also Address of a Minority of the Maryland Ratifying Convention, MARYLAND GAZETTE, May 6, 1788, reprinted in 5 Storing, *supra* note 83, at 97 (expressing the fact that the minority delegates to the Maryland state ratifying convention greatly feared the power of Congress to "subject[] all men able to bear arms to martial law at any moment"); *The Impartial Examiner*, Essay of February 27, 1788, VIRGINIA INDEPENDENT CHRONICLE, Feb. 27, 1788, reprinted in 5 Storing, *supra* note 83, at 181 ("The severity of discipline necessary to be observed [in a standing army] reduces [soldiers] to a degree of slavery . . .").

country's military establishment. Again, attention at the founding was on setting up the military as an institution, and given how daunting a task this proved to be, there was precious little time and energy left over for political fights regarding the rights of the soldiers.⁴³³ Beyond this, the small numbers actually enlisted into the army combined with the fact that most of society held regular soldiers in low regard, meant that a large part of the polity were simply uninterested in protecting the political interests of individual service members.

During the ratification debate defenders of the Constitution sought to assuage the polity's fear of a standing army by claiming that it would seldom be necessary for the country to have a large military force given its distance from the threat presented by the European powers.⁴³⁴ With regards to the number of men enlisted into the military, this claim was an accurate description of the state of affairs when written and it proved to be true as the country moved through the 1790s. In "January of 1784, all that remained of the [national] army . . . was a small corps of artillery and one regiment of infantry, about 600 rank and file under Henry Knox's command."⁴³⁵ Five and a half years later, in August of 1789, President Washington reported to the Congress that there were a total of 840 troops in the service of the United States.⁴³⁶ As we have seen,

⁴³³ See generally Epps, *supra* note 236, at 522.

We can understand why the framers were less than thrilled to be setting out limits on their own power. They were overwhelmingly concerned with the task of governing a new and weak country in a dangerous world. Many had been firebrands in 1776; but now it was 1789, and age cools revolutionary zeal. They were also elected officials of a government, and elected officials tend to see the task of government as being—well—to govern. There is an old saying that to a man with a hammer, everything looks like a nail. The First Congress had a federal structure to build, and they wanted to drive those nails.

Id.

⁴³⁴ See, e.g., THE FEDERALIST NO. 8 (Alexander Hamilton); "An American Citizen" [Tench Coxe], *An Examination of the Constitution of the United States*, INDEPENDENT GAZETTEER, Sept. 26-29, 1788, reprinted in FRIENDS, *supra* note 170, at 473.

⁴³⁵ KOHN, EAGLE, *supra* note 24, at 41.

⁴³⁶ *Military Force in 1789*, *supra* note 310, at 5.

throughout the 1790s the size of the national army expanded and contracted in response to the various threats faced by the country, but it never exceeded more than 6000 men under arms.⁴³⁷ Thus, at the peak of its manpower strength, less than one half of one percent (0.50%) of the able bodied men of the country were serving on active duty in the army.⁴³⁸ The smallness of this force meant that the welfare of regular soldiers held little salience for the vast majority of the population.

To the extent the polity of late 1700s America stopped to reflect upon those who chose to serve in the regular army, their thoughts were not likely to be charitable. In direct contrast to the place of honor held by the general militia, regular soldiers of this period were widely perceived to be the "dregs of society."⁴³⁹ Because pay and rations were set so low by Congress, it was hard for the army to attract anyone who was able to make their livelihood in another way.⁴⁴⁰ As a result, the regular army of the late eighteenth century was not composed of a cross-section of the community; rather, it was made up of "men without land, homes, families, or principles" who "had sold themselves into virtual bondage to the government."⁴⁴¹ More-

⁴³⁷ See, e.g., *supra* note 409 and accompanying text.

⁴³⁸ In the late eighteenth century, men of military age amounted to about a fourth or a fifth part of the whole body of the people. See ADAM SMITH, 2 *THE WEALTH OF NATIONS* 189 (6th ed. 1950). The total population of the United States in 1800 was a little over 5,300,000. DODD, *supra* note 228, at 103. Therefore, the total number of able bodied men available for military service in America in 1799 was between 1,325,000 (5,300,000 x .25) and 1,060,000 (5,300,000 x .20).

⁴³⁹ AMAR, *supra* note 199, at 55; see also *The Impartial Examiner, Essay of Febuary 27, 1788*, VIRGINIA INDEPENDENT CHRONICLE, Feb. 27, 1788, reprinted in 5 Storing, *supra* note 83, at 181 (referring to regular soldiers as "dregs of the people"). Like the prevailing views on the threat of standing armies and the virtue of the militia, the polity's opinion of regular soldiers had its roots in English Whig rhetoric. See, e.g., REID, *supra* note 37, at 105 (noting that regular soldiers in the British army "are generally men who had no character, or perhaps a bad one before they listed in the army") (quoting William Beckford, Commons Debates (Dec. 9, 1754), in 15 *PARLIAMENTARY HISTORY OF ENGLAND FROM THE EARLY PERIOD TO THE YEAR 1803*, at 388-89 (1806-1820)).

⁴⁴⁰ KOHN, EAGLE, *supra* note 24, at 63; see also *id.* at 69 ("At the end of the decade the army was owed nearly \$150,000 in pay and subsistence (the monthly pay of a private was \$7)).

⁴⁴¹ AMAR, *supra* note 199, at 53. The Anti-Federalist who used the pen name

over, unlike men in the general militia who by serving "alongside their families, friends, neighbors, classmates, and fellow parishioners . . . [were] constantly reminded of civil(ian/ized) [sic] norms of conduct," the ties between regular soldiers and the rest of society were further weakened by their full-time service in an isolated military.⁴⁴²

The rootlessness of the average enlisted man contributed to the creation of a force that was characterized all too frequently by low morale, habitual drunkenness, desertion, and mutinous acts.⁴⁴³ The response to this behavior on the part of those in command was to administer ever more harsh methods of discipline in an effort to gain the upper hand.⁴⁴⁴ It was common for corporal punishment in the form of lashes to be employed for many offenses,⁴⁴⁵ and, on occasion, summary executions

"The Federal Farmer," decried giving over military control to "distinct bodies of military men, not having permanent interests and attachments in the community." Letter from The Federal Farmer (Jan. 25, 1798), *reprinted in* 2 Storing, *supra* note 83, at 341; *see also* CUNLIFFE, *supra* note 69, at 103 (emphasis omitted) (In his book *Life of Washington*, author Mason Weems's said of the regular soldiers who had died in the Battle of Fallen Timbers in 1794, "[t]all young fellows, who could easily get their half dollar a day at the healthful and glorious labours of the plough, to go and enlist and rust among the lice and itch of a camp, for four dollars a month, were certainly not worth their country's crying about.").

⁴⁴² AMAR, *supra* note 199, at 53. *See* KOHN, EAGLE, *supra* note 24, at 2-3.

⁴⁴³ KOHN, EAGLE, *supra* note 24, at 69; *see also* *Causes of the Failure of the Expedition Against the Indians, in 1791, Under the Command of Major General St. Clair*, in CLASS V AMERICAN STATE PAPERS: 1 MILITARY AFFAIRS 36, 37 (Walter Lowrie & Walter S. Franklin eds., 1832) (noting that Major General St. Clair had moved his force from Fort Washington to Ludlow's station, in part, to enhance discipline by depriving the men of the means of intoxication); Letter from George Washington to the President of Congress (Sept. 24, 1776), in AMERICAN MILITARY THOUGHT *supra* note 137, at 15 (complaining about "thirty and 40 Soldiers . . . desert[ing] at a time" and the prevalent practice of troops, including officers, plundering civilian property).

⁴⁴⁴ *See, e.g.*, Washington, *supra* note 443, at 14-15 ("For the most atrocious offenses . . . a Man receives no more than 39 Lashes; and these perhaps (thro' the collusion of the Officer who is to see it inflicted), are given in such a manner as to become rather a matter of sport than punishment . . ."). *See generally* The Massachusetts Articles of War of 1775, Preamble, in WILLIAM WINTHROP, MILITARY LAW AND PRECEDENTS 1471 (2d ed. 1920) (referencing the "cruel punishments [that] are usually practiced in Standing Armies").

⁴⁴⁵ *See e.g.*, KOHN, EAGLE, *supra* note 24, at 63. In 1784, then-Lieutenant-Colonel Commandant Josiah Harmar court-martialed four enlisted men and sentenced

would follow the commission of more serious crimes.⁴⁴⁶ The imposition of this harsh discipline tended to brutalize the men in the regular army, thus isolating them even more from civilian society. While some, as detailed above, expressed concern about individual soldiers being treated this way,⁴⁴⁷ many were of the opinion that "[w]retches miserable enough to volunteer as hired guns . . . deserve[d] whatever treatment that they got at the hands of army officers."⁴⁴⁸ Again, a clear difference was drawn between the "wretched" regular soldiers and general militiamen. While the former experienced harsh discipline at the command of unfamiliar officers, the need and ability to discipline the latter was "tempered by the many social, economic, and political linkages that predated military service and that would be renewed thereafter."⁴⁴⁹

At the end of the 1800s it was certainly the case that the military community stood apart from civilian society. An exceedingly small percentage of the able bodied men of the time actually served in the regular army, and those that did were held in low regard by the general population. There were some who recognized that "[t]he political dangers of . . . rel[ying] on military establishments were mirrored by . . . harm to individual citizens,"⁴⁵⁰ but they were in the minority and their comments were predominately aimed at protecting the rights of militia men called into federal service, instead of expressions of concern about the handling of regular soldiers.

each to 100 lashes for destroying private property. *Id.*

⁴⁴⁶ See, e.g., *id.* at 70-71. In attempt to stem the "decimating" effects of desertion on the army at Fort McIntosh in the winter of 1785-1786, Major John Wyllys, post commandant, apprehended three deserters and executed them immediately without trial. *Id.*

⁴⁴⁷ See, e.g., *supra* note 432 and accompanying text.

⁴⁴⁸ AMAR, *supra* note 199, at 54.

⁴⁴⁹ *Id.* at 54-55. Militia officers knew "that, in a variety of ways, they could be called to account back home after the fighting was over." *Id.* at 55.

⁴⁵⁰ Levin, *supra* note 20, at 1048.

CONCLUSION

In word and deed the founding generation explicitly embraced the principles of civilian control over the military and a separation of those civilian powers over that institution. Although it was not nearly as compelling to the polity of late eighteenth century America as these first principles, the need to strike an appropriate balance between military efficiency and individual rights also has its roots in our founding experience. The nascent concern that was present at the founding for individual soldiers has grown into a constitutional imperative as the place that the military establishment occupies in our country has evolved. Today, in order for the military establishment to fit within our constitutional structure it must be subject to civilian control, its management must be divided among the three branches of government, and it must respect the constitutional rights of its individual members.

The strong form of the separate community doctrine violates all three of the constitutional conditions just presented. First, by denying serious judicial review to rights impinging decisions and actions taken solely by the military establishment, the separate community doctrine allows the military to override individual constitutional rights without civilian oversight. Second, by using the separate community doctrine to justify ignoring its own place in our constitutional scheme of separation of powers, the Supreme Court deprives an entire body of jurisprudence of the voice in American law that is best prepared and situated to balance the conflicting interests and values that are a part of the separate community cases. Third, a judicial posture that defers almost entirely to the decisions and actions of the military establishment leaves little room for consideration of any competing individual rights.

Each of the points raised in this brief Conclusion deserve more exploration. This Author has previously examined the creation, development, and present condition of the separate community doctrine and explained how that concept has evolved into a standard of review that leads the Supreme Court to take a largely hands off approach when faced with disputes

between the military establishment and individual service members.⁴⁵¹ This publication, in an effort to lay the groundwork for a more in-depth and intelligent critique of the strong form of the separate community doctrine, has explored the place that the military establishment occupied at the country's founding. Future efforts need to develop a more complete picture of the first principles identified herein, show in greater detail how each of these are compromised by the strong form of the separate community doctrine, and articulate a substitute for that doctrine which will better serve the constitutional imperatives of civilian control, separation of powers, and respect for the rights of individual service members.

⁴⁵¹ Martin, *supra* note 1.

WORKING FOR FREE: IT OUGHT TO BE AGAINST THE (TAX) LAW

*Richard Winchester**

I. INTRODUCTION

Congress often enacts tax laws in order to affect people's behavior in one way or another. However, tax legislation frequently produces outcomes that Congress may neither foresee nor desire. This article describes how the temporary cut in the tax on corporate dividends has produced just such an unanticipated outcome.¹ The law was promoted as a way to reduce economic distortions and undesirable tax planning by corporations.² However, it has also created an opportunity for the rich to improperly avoid tax when they work for a corporation they also own and control.

An individual who owns and works for a corporation has

* Assistant Professor of Law, Thomas Jefferson School of Law; J.D. Yale Law School; A.B., Princeton University. Thanks are in order for Dorothy Brown, Karen Burke, Francine Lipman, Julie Greenberg, Steve Semeraro, Deven Desai, Linda Keller, Anders Kay, and Arne Rosenberg who provided helpful comments when this article was in draft form. This project also received valuable assistance from the work of three students: Amanda Van, Gavin Morrissey, and James Binnal. Lastly, the project would not have been possible without the financial support of the Thomas Jefferson School of Law.

¹ See Pub L. No. 108-27, § 301(a)(1), 117 Stat. 752 (2003).

² The original proposal was to eliminate the tax on corporate dividends so that tax considerations would not affect decisions regarding (1) whether a business is financed by debt or equity, (2) whether a business operates through a corporation or an unincorporated business entity, (3) whether a business distributes its profits to shareholders, (4) whether a corporation repurchases its shares instead of paying dividends to shareholders, and (5) whether a corporation engages in artificial transactions designed solely to minimize the tax on corporate profits. Department of the Treasury, GENERAL EXPLANATIONS OF THE ADMINISTRATION'S FISCAL YEAR 2004 REVENUE PROPOSALS 11 (February 2003).

at least two ways to access the earnings of the business. As a shareholder, he can access the earnings by receiving a dividend on his stock. Alternatively, as an employee, he can withdraw earnings in the form of compensation, such as a salary or bonus. Before the cut in the tax on dividends, such an employee-shareholder almost always had an economic incentive to access the earnings as compensation. However, the cut in the dividend tax reversed the incentives for many employee-shareholders. There is now a substantial economic incentive for high income employee-shareholders to substitute dividends for the compensation they would have otherwise paid themselves. Ironically, certain rich individuals can get richer by working for free.

Although the cut in the dividend tax created an incentive for some employee-shareholders to substitute a dividend for a bonus, the tax cut itself is not the source of the problem. Rather, the tax cut has merely cast a light on the inconsistent and problematic way that federal employment tax rules apply to individuals who work for a business they also own and control. Such employee-owners currently are subject to a different set of rules depending on the legal form used to operate the business. In addition, if the business operates as a partnership, an employee-owner's employment tax liability also depends on whether he is a general partner or a limited partner. To make matters worse, it is not clear what rules apply to an employee-owner of a business operated through a limited liability company.

It is widely acknowledged that the employment tax rules should be made more clear and consistent. However, all past proposals for correcting this problem have not addressed it in a comprehensive way, focusing instead on adjustments that would make the rules apply uniformly only to individuals who work for a business they conduct through an entity other than a taxable corporation.³ Such business forms generally include

³ See proposed regulations at 59 Fed. Reg. 67,253-01 (Dec. 29, 1994). Those proposed rules were withdrawn and replaced with proposed regulations at 62 Fed. Reg. 1702-01 (Jan. 13, 1997). Congress later imposed a moratorium on adminis-

partnerships, S corporations and limited liability companies. These proposals are an improvement over the current situation. However, because they fail to address an advantage enjoyed by individuals who own and work for a business conducted through a taxable corporation, these proposals represent only a partial solution to the problem.

This article will propose a uniform set of employment tax rules that should apply to all individuals who work for a business they own and control. Before doing so, the article will examine the current rules that determine how the earnings of a business are taxed when received by an owner who also works for the business. This will require an examination of two sets of rules. First, there are rules that tax the owner's share of the profits of the business. Second, there are rules that tax amounts paid to the owner as compensation for services rendered to the business. This discussion will reveal how the use of different business entities, different ownership interests and different payout forms can affect what an employee-owner pays as tax and what he keeps after tax. The article will then describe and critique a widely endorsed proposal for eliminating the inconsistencies in the current system of tax rules that apply to employee-owners of non-taxable business entities.

Next, the article shows how a shareholder who works for a solely owned corporation enjoys a significant opportunity to avoid employment tax that other employee-owners do not enjoy. Among other things, the discussion will demonstrate how the temporary dividend tax cut has created an incentive for an employee-shareholder to access the earnings of the business by substituting dividends for any compensation he would otherwise want to receive in the absence of the tax cut. Moreover, the discussion reveals how this outcome has a con-

trative rulemaking in this area. Taxpayer Relief Act of 1997, P.L. 105-34, § 935, 111 Stat. 788 (1997). See also Thomas E. Fritz, *Flowthrough Entities and the Self-Employment Tax: Is it Time for a Uniform Standard?*, 17 VA. TAX REV. 811 (1998); Sheryl Stratton, *ABA/AICPA Have Legislative Fix for LLC Self-Employment Tax Problem*, 84 TAX NOTES 351 (July 19, 1999).

siderable class bias in two respects. First, high income individuals are considerably more likely to be in a position to make this tax-saving substitution. Second, when they do take advantage of it, these high-income employee-shareholders save far more tax dollars than their lower income counterparts would. The article concludes by advancing a proposal that would make employment tax rules apply uniformly to all individuals who work for a business they also own.

II. TAXATION OF BUSINESS PROFITS TO AN EMPLOYEE-OWNER

When an individual works for a business that he also owns, several federal laws may apply to extract a tax on the earnings of the business. The total tax extracted will determine how much the employee-owner has left to spend on personal items unrelated to the business. There are two sets of tax rules to consider. First, there are income taxes that apply. These taxes may be imposed on the employee-owner, the business, or both. Second, there are federal employment taxes that may also come into play to the extent the earnings of the business are treated as the employee-owner's income from labor. The following sections describe the pertinent aspects of each set of rules.

A. Income Taxes on the Profits of a Business

State law recognizes several vehicles through which a business can be conducted, including the sole proprietorship, various forms of the partnership, the limited liability company, and the corporation. Each business form offers a different mix of features that may affect how suitable it may be for any given situation and how attractive it may be to the owners of the business. One factor that the owners of any business are likely to consider is the extent to which the earnings of the business will be subject to the income tax. The rules that apply in any given situation will depend in the first instance on the state law business form used to operate the enterprise. The state law business form will dictate the default tax rules that will govern how the profits of the business are taxed.

However, in most cases, the business can choose whether it wants an alternative set of rules to apply.

The Internal Revenue Code employs two alternative models for taxing the profits of a business. The first is the corporate model, which treats the business entity and its owners as separate and distinct taxpaying units. As a result, the business itself is subject to tax on any profits it makes.⁴ In addition, the owners are subject to tax on any profits that the business actually pays to them as a return on their investment.⁵ The corporate model is the default system of taxation that applies when a business operates through a corporation.⁶

The two-tiered structure embodied by the corporate model stands in contrast to the flow through model of taxation. It does not treat the business entity as a taxpaying unit. Instead, the owners of the business are taxed on their share of the profits of the enterprise, whether or not the business actually pays any of these profits to them. The Internal Revenue Code contains several versions of the flow-through model of taxation. This article will address three.⁷ There are the rules that apply to sole proprietors, there are the rules that apply to partnerships, and there are the provisions of subchapter S.

Federal tax law does not consider a sole proprietor to be a separate business entity.⁸ As a result, any income or loss of the business is merely included in the computation of the owner's individual income tax liability.⁹ Any limited liability company that has only one member is treated as a sole proprietor for federal income tax purposes.¹⁰ However, such a single member limited liability company can choose to be treated as

⁴ See I.R.C. § 11(a) (2000).

⁵ I.R.C. § 61(a)(7).

⁶ Treas. Reg. § 301.7701-2(b)(1) (2006).

⁷ The other versions of the flow-through model address unique industries and situations that are of little relevance to the average self-employed individual. They include the regulated investment company rules, the real estate investment trust rules, and rules that apply to estates and trusts, common trust funds, and qualified electing funds.

⁸ Treas. Reg. § 301.7701-1(a)(2).

⁹ See I.R.C. § 61(a)(2).

¹⁰ Treas. Reg. § 301.7701-3(b)(1)(ii).

a corporation for federal income tax purposes, which would bring the corporate model of tax into play.¹¹ If a business operates through a partnership or a limited liability company that has more than one owner, the business profits are taxed under the partnership version of the flow through model.¹² However, any such business has the option to be treated as a corporation for federal income tax purposes.¹³ When this option is chosen, the corporate model of taxation applies exclusively. If a corporation meets certain eligibility requirements, it can choose to be subject to the flow-through provisions of subchapter S.¹⁴

Unless indicated otherwise, the next several sections use the term corporation to refer to any business entity (other than an S corporation) that is treated as such for federal income tax purposes, regardless of the state law business form through which the enterprise is conducted. Likewise, the article uses the term partnership to refer only to those state law partnerships that have not elected to be treated as a corporation for federal income tax purposes. In addition, the term limited liability company refers to any such company with more than one owner that has not elected to be treated as a corporation for federal income tax purposes. Finally, the term sole proprietor refers to any sole proprietor and any single owner limited liability company that has not elected to be treated as a corporation for federal income tax purposes.

1. Taxation of the Profits of a Corporation

As a general proposition, there are two separate income taxes that apply to the profits of a business conducted through a corporation. First, the corporation itself has to pay an income tax on what it earns.¹⁵ Second, a shareholder must pay an income tax on any after-tax profits that the corporation

¹¹ Treas. Reg. § 301.7701-3(a).

¹² Treas. Reg. § 301.7701-3(b)(1)(i).

¹³ Treas. Reg. § 301.7701-3(a).

¹⁴ I.R.C. §§ 1362(a)(1), 1363(a).

¹⁵ I.R.C. § 11.

pays to him as a dividend.¹⁶ This two-tiered tax structure is one of the hallmarks of the U.S. corporate tax scheme.

The corporate tax applies only to the taxable income of a corporation.¹⁷ Taxable income refers generally to revenues, reduced by the firm's cost of goods sold and certain expenses allowed by law.¹⁸ Among other things, a corporation can deduct amounts paid as compensation to any employee (including an employee-owner) for services rendered to the business.¹⁹ The principal restriction is that the deduction is limited to amounts that are reasonable for the services performed.²⁰ Thus, a corporation's taxable income gets reduced to the extent it pays compensation to its employees, resulting in a lower tax bill.

The corporate tax itself is determined under a system of marginal rates that applies to the corporation's taxable income.

Under this structure, a corporation's taxable income consists of several different layers of income, each of which is taxed at a different rate. The first layer consists of all income up to \$50,000, which is taxed at 15%.²¹ The second layer consists of all income over \$50,000 and up to \$75,000, which is taxed at 25%.²² Each successive layer covers a higher range of taxable income, starting where the preceding layer left off. And the statute prescribes a different rate that applies to each of these layers. The marginal rates range from a low of 15% to a high of 39%. The following table summarizes the range of taxable income covered by each layer and the tax rate that applies to each layer.²³

¹⁶ I.R.C. § 61(a)(7).

¹⁷ See I.R.C. § 11(a).

¹⁸ I.R.C. § 63(a).

¹⁹ I.R.C. § 162(a).

²⁰ I.R.C. § 162(a)(1).

²¹ I.R.C. § 11(b)(1)(A).

²² I.R.C. § 11(b)(1)(B).

²³ See I.R.C. § 11(b)(1).

Corporate Tax Rates
Taxable Income

Over	Up To	Tax Rate
\$0	\$50,000	15%
\$50,000	\$75,000	25%
\$75,000	\$100,000	34%
\$100,000	\$335,000	39%
\$335,000	\$10,000,000	34%
\$10,000,000	\$15,000,000	35%
\$15,000,000	\$18,333,333	38%
\$18,333,333	unlimited	35%

Thus, if a corporation has \$150,000 of taxable income, that income will consist of four layers. The first \$50,000 will be taxed at 15%, the next \$25,000 will be taxed at 25%, the next \$25,000 will be taxed at 34%, and the last \$50,000 will be taxed at 39%.

Any profits that remain after the corporate tax has been extracted will be subject to tax again if they are paid to the shareholder as a dividend.²⁴ The amount the shareholder pays in tax will depend on several factors. The first consideration is the year when the dividends are paid. Different rules apply depending on the year the dividend is paid.

Dividends paid before 2003 and after 2010 are taxed under the same rules that apply to all income of the shareholder other than gains from the sale of capital assets held for over a year.²⁵ Such ordinary income is subject to tax under a system of marginal rates, similar to the system of marginal rates that applies to corporations.²⁶ Thus, the tax only applies to the extent the shareholder has taxable income. The taxable income

²⁴ I.R.C. § 61(a)(7).

²⁵ Cf. I.R.C. § 1(h).

²⁶ See I.R.C. § 1.

of an individual generally includes all income (other than those items that are expressly exempt from tax) reduced by any deductions, exclusions and exemptions allowed by law.²⁷ Dividends received on corporate stock are included in the computation of taxable income.²⁸

An individual's taxable income consists of several different layers of income, each of which is taxed at a different rate. The personal income tax imposes a set of marginal tax rates that is different from the rates imposed under the corporate income tax. In addition, each marginal rate applies to a different range of income depending on an individual's filing status. There are four categories into which an individual tax return can fall: Married Individual's Filing a Joint Return, Heads of Households, Unmarried Individuals, and Married Individuals Filing a Separate Return.²⁹ In each case, there are six marginal tax rates ranging from 10% to 35%. The following table summarizes the six ranges of taxable income and the tax rate that applies to each range in the case of married individuals who file a joint return in 2006.³⁰

²⁷ See I.R.C. § 63.

²⁸ I.R.C. § 61(a)(7).

²⁹ See I.R.C. § 1.

³⁰ I.R.C. § 1(a), as modified by Rev. Proc. 2005-70, 2005-47 I.R.B. 979 (2005), to reflect inflation adjustments required by law.

**2006 Income Tax Rates
Married Individuals Filing Jointly**

Taxable Income Over	Taxable Income Up To	Tax Rate
\$0	\$15,100	10%
\$15,100	\$61,300	15%
\$61,300	\$123,700	25%
\$123,700	\$188,450	28%
\$188,450	\$336,550	33%
\$336,550	unlimited	35%

Thus, a married couple with \$100,000 of combined taxable income in 2006 will be taxed at 10% on the first \$15,100. A 15% tax will apply to the next \$46,200 of taxable income. And a 25% tax will apply to the last \$38,700 of taxable income. Those six separate rate categories are in effect through 2010. After 2010, the schedule will be replaced with the one that was in effect from 1993 to 2000.³¹ That schedule contained the following five marginal rates: 15%, 28%, 31%, 36% and 39.6%.³² There were transitory schedules that applied in 2001 and 2002, each one containing six marginal rates. For 2001, the marginal rates were 10%, 15%, 27.5%, 30.5%, 35.5% and 39.1%. For 2002, the marginal rates were 10%, 15%, 27%, 30%, 35% and 38.6%.³³

If a shareholder receives a corporate dividend after 2003 and before 2008, the dividend can be taxed in one of two ways, depending on how long the shareholder owned the stock in the dividend paying corporation. If the shareholder owned the stock for less than sixty-one days, the dividend will comprise part of the shareholder's ordinary income, making it subject to

³¹ Pub. L. No. 107-16, § 901(a)(91), 115 Stat. 38 (2001); Pub. L. No. 108-27, § 107, 117 Stat. 752 (2003).

³² I.R.C. § 1.

³³ I.R.C. § 1(i), as amended by Pub. L. No. 108-27, § 105(a).

tax under the marginal rates just described.³⁴ However, in most cases where the shareholder owned the stock for at least sixty-one days, dividends paid on the stock will be taxed at the same rate that applies to gains from the sale of stock and other capital assets held for over one year.³⁵ That rate varies depending on the top marginal tax rate that applies to the shareholder's ordinary income. If that rate is 25% or higher, the dividends are taxed at 15%.³⁶ If that rate is below 25%, the dividends are taxed at 5%.³⁷

If a shareholder receives a corporate dividend in 2008 through 2010, the sixty-one-day minimum holding still applies to determine whether the dividend will comprise part of the shareholder's ordinary income or if it will be taxed as a long term capital gain. In addition, the tax imposed in the latter situation will still vary depending on the marginal tax rate that applies to the shareholder's ordinary income. The only difference is that the rates will change. The tax will be 15% if the shareholder's ordinary income is subject to tax at the marginal rate of 25% or higher. However, the dividend will be exempt from tax if the shareholder's ordinary income is subject to tax at a marginal rate below 25%.³⁸ The following table summarizes by year how dividends are taxed to individual shareholders.³⁹

³⁴ Cf. I.R.C. § 1(h)(11)(B)(iii)(I).

³⁵ I.R.C. § 1(h)(11)(A)), as amended by Pub. L. No. 108-27, § 302(a) (2003). I.R.C. § 1(h)(3)(B)), as amended by Pub. L. No. 108-27, § 302(e)(1). See Pub. L. No. 108-22, § 303, 117 Stat. 696 (2003) (providing sunset for this provision after 2008).

³⁶ I.R.C. § 1(h)(1)(C).

³⁷ I.R.C. § 1(h)(1)(B).

³⁸ I.R.C. § 1(h)(1)(B) (2006), as amended by the Tax Increase Prevention and Reconciliation Act of 2005, Pub. L. No. 109-222, § 102, 120 Stat. 345 (2006).

³⁹ Different rules apply when dividends are paid to shareholders that are corporations. See I.R.C. §§ 243-246A.

**Tax on Corporate Dividends
Received by Individuals**

Year Dividend Received

Minimum Holding Period Met	Before 2001	2001	2002	2003-2007	2008	2009-2010	After 2010
Yes	N/A	N/A	N/A	15% or 5%	15% or 0%	N/A	N/A
No	Ordinary Income: 15% 28% 31% 36% 39.6%	Ordinary Income: 10% 15% 27.5% 30.5% 35.6% 39.5%	Ordinary Income: 10% 15% 27% 30% 35% 38.6%	Ordinary Income: 10% 15% 25% 28% 33% 35%	Ordinary Income: 10% 15% 25% 28% 33% 35%	Ordinary Income: 10% 15% 25% 28% 33% 35%	Ordinary Income: 15% 28% 31% 36% 39.6%

Although the rules subject the profits of a business to tax at the corporate level and also at the shareholder level, there are many situations in which only one of the two taxes will apply. For instance, the shareholders will not have to pay tax on any profits that are not actually paid to them as dividends. In such a case, only the corporation will be subject to tax on the earnings. Alternatively, only a shareholder will be taxable on amounts paid to him as reasonable compensation for services rendered to the firm. The corporation will pay no tax on these amounts because they are a deductible item that reduces the corporation's taxable income.

Because shareholders of (primarily closely held) corporations have an incentive to employ such strategies to avoid tax, there are rules designed to either penalize or outlaw such behavior. A practice of not paying dividends to shareholders could trigger the accumulated earnings tax. The Internal Revenue Service is authorized to assess this penalty tax when it determines that the corporation has accumulated profits beyond the reasonable needs of the business.⁴⁰ The tax is generally computed at a rate that corresponds to the rate that would apply to any dividends paid to a shareholder. Thus, the

⁴⁰ I.R.C. §§ 531, 532(a), 533(a).

current rate is 15%.⁴¹

Aside from the accumulated earnings penalty tax, the personal holding company tax would apply if two conditions are met. First, at least 60% of the corporation's gross income must come from certain passive sources, like interest and dividends.⁴² Second, five or fewer individuals must own over half of the stock in the corporation.⁴³ If these two conditions are met in any given year, a penalty tax is imposed on virtually all taxable income of the corporation from that year, other than amounts distributed to shareholders as a dividend.⁴⁴ Unlike the accumulated earnings tax, the personal holding company tax applies in a mechanical fashion. It cannot be avoided by showing the absence of an intention to evade tax.⁴⁵

Like the accumulated earnings tax, the penalty is computed at a rate that generally corresponds to the rate that would apply to dividends paid to a shareholder. Thus, the current penalty rate is 15%. After 2010, when the temporary tax cut on dividends expires, the rate will be adjusted to reflect the highest marginal tax rate that applies to individuals. That rate currently stands at 35%, but it is due to revert to 39.6% after 2010.

In cases where the corporation uses its profits to pay a shareholder compensation for services rendered to the business, there is only one restriction on the firm's ability to do so: the amount must be reasonable under all circumstances.⁴⁶ The absence of a bright line rule operates as an invitation to those who would exploit the ambiguity to their advantage, knowing that the chances of being caught and penalized are slim. Because unreasonable compensation is difficult to detect,

⁴¹ I.R.C. § 531.

⁴² I.R.C. § 542(a)(1).

⁴³ I.R.C. § 542(a)(2).

⁴⁴ I.R.C. §§ 541, 545.

⁴⁵ *Cf.* I.R.C. §§ 532(a), 533. When the personal holding company tax applies, the accumulated earnings tax will not. I.R.C. § 532(b)(1).

⁴⁶ I.R.C. § 162(a)(1); *Treas. Reg.* § 1.162-7(a)(3).

many such overstated deductions go unchallenged.⁴⁷

2. Taxation of the Income of an S Corporation

Federal tax law permits any corporation that satisfies certain eligibility requirements to elect to be treated as a small business corporation for federal tax purposes.⁴⁸ As a result, the corporation would be subject to the rules of subchapter S of the Internal Revenue Code. Under subchapter S, the corporation does not pay any corporate tax on the income of the business.⁴⁹ Instead, the business profits are taken into account by each shareholder to determine their individual income tax liabilities.⁵⁰ The amount that any shareholder takes into account is referred to as the shareholder's pro rata share and reflects his percentage interest in the corporation.⁵¹ Thus, an individual who owns half the stock in an S corporation in any given year is taxable on half of any income (or loss) generated by the corporation in that year.

The actual tax that an S corporation shareholder pays on any income allocated to him will depend on the source of the income.⁵² Thus, any item of income that the corporation derives from a tax exempt source will be tax exempt to the

⁴⁷ See *infra* text accompanying notes 176-184.

⁴⁸ Under current law, an S corporation can have no more than 100 shareholders, all of whom must be individuals (other than nonresident aliens), certain trusts or estates. I.R.C. § 1361(b)(1)(A)-(C). In addition, there can only be one class of stock in the corporation. I.R.C. § 1361(b)(1)(D). Finally, the S election is not available to certain financial institutions, insurance companies, domestic international sales corporations, or corporations that have made an election under § 936. I.R.C. § 1361(b)(2).

⁴⁹ I.R.C. § 1363(a). There are two very narrow exceptions to this general rule. In certain cases where an S election has not been in effect throughout the life of the corporation, an S corporation may be liable for two different taxes. A tax on built-in gains could apply if an S corporation realizes a gain on the sale of an appreciated asset it owned at the time the S election took effect. I.R.C. § 1374. Second, if the S election took effect before the time the corporation distributed any accumulated profits to its shareholders, a tax on excess passive investment income could apply until the corporation actually distributes those accumulated profits to its shareholders. I.R.C. § 1375.

⁵⁰ I.R.C. § 1366(a).

⁵¹ I.R.C. § 1377(a)(1).

⁵² I.R.C. § 1366(b).

shareholder. Similarly, any item of income that the corporation derives from the sale of a capital asset that the corporation held for over a year will be taxed at the rate that applies to long term capital gains.⁵³ Meanwhile, any item of income that the corporation derives from a source other than one that is subject to a special rate of tax is treated as ordinary income, making it subject to tax at the marginal rates that apply to all of the shareholder's ordinary income.⁵⁴ An S corporation is commonly referred to as a flow-through entity precisely because any income generated by the corporation retains its character when allocated to the shareholders as pro rata shares. Moreover, this income is taxed to the shareholders in the year earned by the corporation, not the year received from the corporation.

A shareholder in an S corporation is taxable on amounts he actually receives from the corporation depending on the reason for the payment. To the extent the payment is a distribution of the business profits previously allocated to the shareholder, the payment will not be taxable to him.⁵⁵ (The tax was already imposed when the corporation allocated its income to the shareholder as a pro rata share.) However, if the shareholder receives a payment that constitutes compensation for services rendered to the corporation, the payment will be taxable as income to him, just as compensation from any other source would.⁵⁶ Such compensation will also be a deductible expense to the corporation, provided the amount is reasonable for the services rendered.⁵⁷ Like any other deductible expense, the payment will reduce the amount of business profits that are left to be allocated to the shareholders as taxable pro rata shares.⁵⁸ The net effect is that the income tax of a shareholder in an S corporation is based partly on amounts that the corporation pays to him as compensation, and partly

⁵³ I.R.C. § 1(h).

⁵⁴ See *supra* discussion accompanying notes 26-30.

⁵⁵ I.R.C. § 1368(b)(1).

⁵⁶ I.R.C. § 61(a)(1).

⁵⁷ I.R.C. § 162(a)(1).

⁵⁸ I.R.C. §§ 1363(b); 1366(a), (c).

on the profits (net of compensation expense) the corporation allocates to him as a pro rata share.

3. Taxation of the Income of a Partnership or Multi-Member Limited Liability Company

Subchapter K contains the rules that determine the extent to which the profits of a business are taxed when the business is conducted as a partnership or limited liability company that has more than one member. As discussed above, both such business entities are treated as a partnership for federal income tax purposes.⁵⁹ As a general rule, the profits of a business conducted through a partnership are not taxed at the partnership level.⁶⁰ Instead, the partners are taxed on their respective shares of the taxable profits of the business.⁶¹ In any given tax year, a partner will be allocated a share of the partnership's business profits and other items.⁶² Such an allocation is referred to as the partner's distributive share.⁶³ To the extent a partnership has made an allocation of profits to a partner, the partner will not pay tax on those profits when he actually receives them in the form of a distribution from the partnership.⁶⁴ Thus, if a 50-50 partnership makes \$100,000 in taxable income in one year, each partner will be allocated and taxable on \$50,000. However, neither partner will be subject to tax when they actually receive a distribution of that money from the partnership.

Like an S corporation, a partnership derives its flow-through qualities from the fact that only the partners, not the business, are taxable on the profits of the business.⁶⁵ In addi-

⁵⁹ See *supra* discussion accompanying note 12.

⁶⁰ I.R.C. § 701.

⁶¹ *Id.*

⁶² I.R.C. § 702.

⁶³ I.R.C. § 702(a).

⁶⁴ I.R.C. § 731(a)(1). The statute specifically permits a partner to receive a distribution tax free up to the partner's basis in the partnership. A partner acquires (or loses) basis as a result of an allocation of profits (or losses) to the partner in the form of distributive shares. I.R.C. § 705(a).

⁶⁵ I.R.C. § 701.

tion, a partner's actual tax liability will depend in part on the source of any income item allocated to him.⁶⁶ Thus, if the partnership derives income from a tax exempt source, the partners will pay no tax on their respective allocations of that item. Similarly, any item of income that the partnership derives from the sale of a capital asset that the partnership held for over a year will be taxable to the partner at the rate that applies to long term capital gains.⁶⁷ Meanwhile, any item of income that the partnership derives from a source other than one that is subject to a special rate of tax is treated as ordinary income, making it subject to tax at the marginal rates that apply to all of the partner's ordinary income.⁶⁸

A partner is taxable on amounts he actually receives from a partnership depending on the reason for the payment. To the extent the payment is a distribution of the business profits previously allocated to the partner, the payment generally will not be taxable to the partner.⁶⁹ (The tax was already imposed when the partnership allocated its income to the partner as a distributive share.) However, if the partner receives a payment that constitutes compensation for services rendered to the partnership, the payment will be taxable as income to the partner, just as compensation from any other source would.⁷⁰ Referred to as a form of "guaranteed payment," such compensation paid to a partner will also be a deductible expense to the partnership, provided it is reasonable for the services rendered.⁷¹ Like any other deductible expense, the payment will reduce the amount of business profits that are allocated to

⁶⁶ I.R.C. § 702(b).

⁶⁷ I.R.C. § 1(h).

⁶⁸ See *supra* discussion accompanying notes 26-30.

⁶⁹ I.R.C. § 731(a)(1).

⁷⁰ I.R.C. § 61(a)(1). The payment is taxable income to the partner only if the amount of the payment does not depend on the income of the partnership. I.R.C. § 707(c). Otherwise, the transaction is treated as an allocation of partnership profits to the partner, followed by an actual payment of those profits to the partner. Cf. I.R.C. § 707(a)(2)(A).

⁷¹ I.R.C. § 162(a)(1). The deduction is only available if the amount of compensation does not depend on the income of the partnership. I.R.C. § 707(c).

the partners as taxable distributive shares.⁷² The net effect is that the income tax of a partner will be based partly on amounts that the partnership pays to him as a guaranteed payment of compensation, and partly on the profits (net of compensation expense) the partnership allocates to him as a distributive share.

Although the rules that apply to partnerships and S corporations are similar in many ways, the partnership rules generally allow for more flexibility. First, there are several restrictions on the number and the type of shareholders that a corporation can have in order to be eligible to convert to (and remain) an S corporation.⁷³ These restrictions stand in contrast to the absence of any limitations on the types and number of investors that can be partners in a partnership. In addition, in an S corporation, all items of income, loss and deduction, and credit must be allocated among the shareholders each year pro rata, based on the number of shares they own and the length of time they owned the shares.⁷⁴ By contrast, in a partnership, virtually any allocation that the partners choose will be respected as long as the allocation tracks the economic relationship between the partners.⁷⁵ This generally means that the allocations of partnership income and other items control what the individual partners are entitled to receive in the form of actual payouts.⁷⁶ Moreover, because an S corporation can only have one class of stock, it cannot single out any individual shareholder for the payment of dividends.⁷⁷ Rather, any distribution of business profits must be shared by all shareholders on the basis of the number of shares owned. By contrast, if a partnership wants to make a distribution to just one partner, it can do so. Similarly, if it wants to make a distribution to several partners, there is no requirement that the distribution be allocated among its re-

⁷² I.R.C. §§ 707(c), 703(a).

⁷³ See *supra* note 48.

⁷⁴ I.R.C. §§ 1366(a), 1377(a)(1).

⁷⁵ I.R.C. § 704(b).

⁷⁶ Treas. Reg. § 1.704-1(b)(2)(ii)(b) (2006).

⁷⁷ I.R.C. § 1361(b)(1)(D).

cipients in any particular way.

B. Federal Taxes on an Employee-Owner's Income from Labor

There are two sets of federal employment tax statutes that may apply to an individual who works for a business he also owns. The first is the Federal Insurance Contribution Act (FICA), which imposes a tax that is commonly referred to as the social security tax. The second is the Self-Employment Contribution Act ("SECA"), which imposes a tax that is often referred to as the self-employment tax. The amounts collected under both acts are earmarked for funding social security and Medicare benefits.⁷⁸ The two acts are mutually exclusive so that only one set of rules will ever apply to any given dollar of earnings. Thus, when the FICA rules apply, the SECA rules will not, and vice versa.

The statutes are intended to impose a tax on income from labor, as opposed to any returns on capital. As a result, each statute attempts to define the tax base in a way that isolates such labor income.⁷⁹ However, because it can be difficult to distinguish such income from what an employee-owner receives as a return on capital, the line that separates the two can often appear arbitrary.⁸⁰

1. FICA

The tax imposed by FICA has two components. The first is the old-age, survivors, and disability insurance component, often referred to as OASDI. It is a 12.4% levy on amounts that constitute "wages from employment."⁸¹ One half of the tax is deducted from the employee's compensation.⁸² The employer

⁷⁸ 42 U.S.C. § 401 (2000).

⁷⁹ Patricia E. Dilley, *Breaking the Glass Slipper—Reflections on the Self-Employment Tax*, 54 TAX LAW. 65 (2000).

⁸⁰ Lester B. Snyder, *Taxation with An Attitude: Can We Rationalize the Distinction Between "Earned" and "Unearned" Income?*, 18 VA. TAX REV. 241 (1998).

⁸¹ I.R.C. §§ 3101(a), 3111(a).

⁸² I.R.C. § 3102(a).

pays the other half.⁸³ This component of the FICA tax is earmarked to cover social security benefits. There is a limit on the amount of wages that can be taxed.⁸⁴ Referred to as the contribution and benefit base, this limit is \$94,200 for 2006.⁸⁵ Thus, any wages from employment beyond that limit are exempt from the FICA-OASDI tax. The contribution and benefit base is adjusted each year to reflect increases in average wages of the U.S. economy.⁸⁶

The second component of the FICA tax is the hospital insurance component. It is a 2.9% levy on an individual's "wages from employment." As with the OASDI component, one half of this tax is deducted from the employee's compensation, while the employer pays the other half. However, unlike the OASDI component, there is no limit on the amount of wages from employment that is subject to the tax. Thus, the hospital insurance tax applies to all amounts that qualify as wages from employment, even amounts that exceed the OASDI contribution and benefit base. The hospital insurance component of the FICA tax is earmarked to cover Medicare benefits.

The FICA tax will apply to a self-employed individual only when that person operates the business through either a C corporation or an S corporation. In those instances, only amounts that the corporation pays to the employee-owner as remuneration for employment count as wages from employment.⁸⁷ Thus, only those amounts are subject to the FICA tax. The individual's share of any other profits of the business simply is not subject to the FICA tax, even if it could be considered the product of the employee-owner's labor. As a result, earnings that the corporation retains are not subject to the FICA tax. Nor are amounts paid to an employee-shareholder as a dividend. By defining the tax base in this way, FICA presents the opportunity for individuals to manage or control

⁸³ I.R.C. § 3111(a).

⁸⁴ I.R.C. § 3121(a)(1).

⁸⁵ I.R.S. Notice 2005-85, 2005-46 I.R.B. 961.

⁸⁶ 42 U.S.C. § 430 (2000).

⁸⁷ I.R.C. § 3121(a).

their employment tax liability when they own and control a business conducted through an S corporation or a C corporation. In such cases, the individual can determine whether compensation is paid, when it gets paid, and how much is paid. By exercising this power, the individual necessarily controls whether he must pay the FICA tax, when he must pay the FICA tax, and how much tax he must pay.

A limited liability company that is treated as a corporation enjoys additional tax planning opportunities. Because shares in a state law corporation belong to designated classes, all owners of shares in a given class must share in any distribution paid to one class member; the corporation cannot single out an individual shareholder to receive a dividend distribution. No such restriction applies to a limited liability company. Thus, the company is entirely free to single out one of its members for a distribution. Similarly, the company could make a distribution to several members and not be obligated to allocate the payment in any particular way. This flexibility presents the opportunity for an employee-member to receive a distribution as disguised compensation for services rendered to a company, potentially avoiding the member's employment tax liability.

2. SECA

The SECA tax operates as the FICA tax counterpart for self-employed individuals. Accordingly, like the FICA tax, the SECA tax has two components. The first component is a 12.4% tax earmarked to finance social security benefits. Its counterpart is the OASDI component of the FICA tax. The second component is a 2.9% tax earmarked to fund the Medicare insurance program.⁸⁸

The contribution and benefit base that applies to the OASDI component of the FICA tax also applies to the OASDI component of the SECA tax. Thus, the OASDI tax applies to no more than \$94,200 in 2006. The SECA and FICA statutes

⁸⁸ I.R.C. § 1401(b).

are designed so that the OASDI component of the taxes will never apply to more than the FICA contribution and benefit base in effect for any year.⁸⁹ Thus, if an individual has \$100,000 of wages from employment in 2006, the FICA-OASDI tax would apply to the first \$94,200, leaving no portion of any self-employment income to be taxed under SECA. Conversely, if an individual has no wages from employment in 2006, there would be nothing to tax under FICA, while the SECA-OASDI tax would apply to up to \$94,200 of any income the individual may have from self-employment. If, however, an individual has \$40,000 of wages from employment in 2006, the entire amount would be subject to the FICA-OASDI tax, while up to \$54,200 of self-employment income would be subject to the SECA-OASDI tax, resulting in a tax on no more than the \$94,200 contribution and benefit base in effect for the year. By operating in this way, the rules ensure that anyone whose income includes both wages from employment and income from self-employment will never be at a disadvantage to someone who does not have income from both sources.⁹⁰

Both components of the SECA tax apply to an individual's "income from self-employment."⁹¹ The term does not include any amounts that are subject to the FICA tax.⁹² In addition, in order to count as income from self employment, an item must qualify as net earnings from self-employment ("NESE").⁹³ What counts as NESE for a self-employed indi-

⁸⁹ See I.R.C. § 1402(b).

⁹⁰ The fact that the OASDI component of the employment tax does not apply to amounts in excess of the annually contribution and benefit base distinguishes it from the generally progressive way in which the federal income tax operates. The income tax applies only to the extent an individual has income that exceeds certain amounts that are either excluded, exempt, or deducted from gross income. I.R.C. § 63. The two federal employment taxes have been criticized for being regressive. *E.g.*, Deborah A. Geier, *Integrating the Tax Burdens of the Federal Income and Payroll Taxes on Labor Income*, 22 VA. TAX REV. 1 (2002).

⁹¹ I.R.C. §§ 1401(a)-(b).

⁹² I.R.C. § 1402(b)(1).

⁹³ I.R.C. § 1402(b). A taxpayer is allowed a deduction in computing net earnings from self-employment. The deduction is equal to one-half of the taxpayer's OASDI and HI tax, determined before taking this deduction into account. I.R.C. § 1402(a)(12).

vidual will depend on the kind of legal entity used to conduct the business enterprise. It will also depend on the kind of ownership interest the individual may have in the business. However, in no event will the SECA tax ever apply to amounts generated by a self-employed individual who conducts his business through either a C corporation or an S corporation. In those cases, the payments they receive as compensation will be subject to the FICA tax.⁹⁴ Any dividends received by a shareholder in a C corporation are expressly excluded from the reach of the SECA tax.⁹⁵ Furthermore, SECA has no provision that would count as part of the tax base of an individual's pro rata share from an S corporation.

Because the S corporation is a flow-through entity, one would expect that the SECA rules would control to determine the employment tax liability of any shareholder, just as they do to the owner of all other flow-through business entities. The fact that it does not is largely a relic of a bygone era. When the self-employment tax was enacted, the S corporation did not exist, so the tax base could not be defined by reference to amounts earned through such a business. Furthermore, when subchapter S was adopted, a shareholder's pro rata share was treated as a dividend.⁹⁶ The SECA statute expressly states that net earnings from self-employment do not include dividends.⁹⁷ However, subchapter S was later revised to modify the tax character of an S corporation's pro rata share. Today, a pro rata share is no longer regarded as a dividend.

⁹⁴ See *supra* discussion accompanying note 87. See also I.R.C. § 1402(c)(2).

⁹⁵ I.R.C. § 1402(a)(2). The Internal Revenue Service has had limited success convincing courts to treat dividends as disguised remuneration for services rendered, causing such amounts to be subject to the FICA tax. *E.g.*, *Nu-Look Design, Inc. v. Comm'r*, 356 F.3d 290 (3d Cir. 2004); *Spicer Accounting, Inc. v. United States*, 918 F.2d 90 (9th Cir. 1990); *Radtke v. United States*, 712 F. Supp. 143 (E.D. Wis. 1989). See also Burgess J.W. Raby & William L. Raby, *Shareholder Compensation: How Low Can You Go?*, TAX NOTES TODAY, June 13, 1996, 96 TNT 116-62.

⁹⁶ I.R.C. § 1373(b) (1959).

⁹⁷ I.R.C. § 1402(a)(2). In addition, the Internal Revenue Service concluded that such an item did not count as part of the shareholder's net earnings from self-employment. Rev. Rul. 59-221, 1959-1 C.B. 225.

Instead, the individual items of S corporation taxable income flow through to the shareholders, retaining their character in the hands of the shareholder.⁹⁸ Even though this made a shareholder's pro rata share virtually identical to a partner's distributive share, Congress never updated the self-employment statute to establish parity in the way the law applies to the two situations. Thus, today the statute does not define net earnings from self-employment to include an S corporation shareholder's pro rata share, while it does expressly include a partner's distributive share of partnership income as such.⁹⁹ Clearly, Congress could update the law if it wanted. It has been suggested that Congress has not done so partly because it views the separate existence of the S corporation as a sufficient basis for treating pro rata share allocations as investment income, not income from labor.¹⁰⁰

a. Partner in a Partnership

The employment tax base of a partner in a partnership depends on whether the partner is a general partner or a limited partner. If a partner is a general partner, the self-employment tax will apply to the partner's distributive share allocation of partnership income.¹⁰¹ The tax will also apply to any guaranteed payment the partner receives, whether for the use of capital or for the performance of services.¹⁰² For a lim-

⁹⁸ I.R.C. § 1366(b), as amended by Subchapter S Revision Act of 1982, P.L. No. 97,354, 96 Stat. 1669 (1982).

⁹⁹ I.R.C. § 1402(a).

¹⁰⁰ Fritz, *supra* note 3, at 825.

¹⁰¹ I.R.C. § 1402(a). Certain adjustments are made to the partner's distributive share to determine the amount that is subject to the self-employment tax. The adjustments generally prevent the tax from applying to certain passive items of income that do not represent income from labor. Thus, in computing the self-employment income of a partner, the distributive share is adjusted to exclude, among other things, interest and dividends, and gains and losses from the sale of capital assets. I.R.C. §§ 1402(a)(2)-(3).

¹⁰² Treas. Reg. § 1.1402(a)-1(b). The regulation predates a 1977 amendment that redefined what counts as self-employment income to a partner. I.R.C. § 1402(a)(13), added by the Social Security Amendments of 1977, Pub. L. No. 95-216, § 313(b), 91 Stat. 1509 (1977). (This paragraph was originally added as paragraph 12. However, Pub. L. No. 98-21, § 124(c)(2), 97 Stat. 65, re-designated

ited partner, the self-employment tax applies only to the guaranteed payments received for the performance of services.¹⁰³ It does not apply to the partner's distributive share of income of the partnership.¹⁰⁴ If a partner owns both a general partnership interest and a limited partnership interest, the self-employment tax applies to that portion of the partner's distributive share associated with the general partnership interest only.¹⁰⁵

There are no provisions in the self-employment tax statute or regulations that specify what distinguishes a limited partner from a general partner for purposes of the statute.¹⁰⁶ This stands in contrast to the standard articulated in the Revised Uniform Partnership Act. Under those rules, a limited partner is not liable for the debts and obligations of the partnership, while a general partner is. Moreover, a limited partner risks losing the protection of the limited liability if he participates in the control of the partnership.¹⁰⁷ Thus, under current law, a partner's exposure for the self-employment tax is purely a matter of the nature of the interest the partner owns in the partnership.

One might expect that amounts received by a partner in exchange for the performance of services would count as wages from employment for FICA purposes. However, the legislative history indicates that Congress expected that it would not be appropriate to treat the partnership as a separate tax-paying entity (as opposed to an extension of the partner) in

paragraph 12 as paragraph 13.) The change only affected what counts as self-employment income to a limited partner. The legislative history does not elaborate on the intended scope of the change. *See* H.R. REP. NO. 95-8702 (1977). Thus, it appears that general partners remain subject to employment tax on guaranteed payments received both for services performed and for the use of capital.

¹⁰³ I.R.C. § 1402(a)(13).

¹⁰⁴ I.R.C. § 1402(a)(13).

¹⁰⁵ Prop. Treas. Reg. § 1.1402(a)-2(h).

¹⁰⁶ However, there are proposed regulations which would consider the degree to which a limited partner participates in the operations of the partnership. Prop. Treas. Reg. § 1.1402(a)-2(h)(2). Congress acted in 1997 to prohibit the Internal Revenue Service from finalizing these regulations. Taxpayer Relief Act of 1997, Pub L. No. 105-34, § 935, 111 Stat. 787 (1997).

¹⁰⁷ Revised Uniform Partnership Act § 303(a).

certain situations.¹⁰⁸ In addition, the Internal Revenue Service long ago concluded that it is inappropriate to treat a partnership as an employer of one of its members.¹⁰⁹ As a result, payments that are considered to be made by the partnership to a partner who is not acting in his capacity as a partner will not count as wages that are subject to the FICA tax. Instead, the amounts are treated as self-employment income to the partner.¹¹⁰

b. Member of a Multi-Member Limited Liability Company

Because a multi-member limited liability company is treated as a partnership for federal income tax purposes (absent an election to be treated as a corporation), any member is treated as a partner in a partnership for purposes of the self-employment tax. However, applying the partnership rules to a limited liability company does not produce any clear answers.

A limited liability company has only one category of member, while a partner in a partnership can participate as either a general or a limited partner, depending on the level of liability protection the owner desires and the extent to which that person intends to participate in the business of the partnership. In a limited liability company, each member enjoys the same level of protection from liability for the debts and obligations of the business; it is limited to the member's investment in the company.¹¹¹ In addition, a member does not jeopardize the limit on liability by participating in the business of the

¹⁰⁸ The 1954 Conference Report includes the following language:

[Section 707(a) provides] for the use of the "entity" approach in the treatment of the transactions between a partner and a partnership No inference is intended, however, that a partnership is to be considered a separate entity for the purpose of applying other provisions of the internal revenue laws if the concept of the partnership as a collection of individuals is more appropriate for such provisions.

H.R. REP. NO. 83-2543 (1954).

¹⁰⁹ Rev. Rul. 69-184, 1969-1 C.B. 256.

¹¹⁰ *Id.*

¹¹¹ Uniform Limited Liability Company Act §§ 303, 402(a) (1996) [hereinafter ULLCA].

company.¹¹²

SECA applies a different set of rules depending on whether the individual is either a general or a limited partner in the partnership. However, a limited liability company draws no such distinction between its members. One could assert a reasoned basis for applying either rule. For example, it would seem appropriate to treat a member as equivalent to a general partner since all members are in a position to participate in the operations of the company.¹¹³ On the other hand, one could argue that the limited partnership rules should apply on the grounds that a member enjoys limited liability from the debts and obligations of the business, the hallmark of a limited partner's status as such.¹¹⁴

Understandably, the absence of a clear rule has been an invitation for some to take the position that a member must comply with the rules that apply to limited partners when doing so works to their advantage.¹¹⁵ Taking that position minimizes the member's employment tax liability because the member's net earnings from self-employment would consist solely of amounts received from the company in exchange for services the member performed for the company; no part of the member's allocation of business profits would be included in the employment tax base.

c. Sole Proprietor

If a single individual does not create or organize a formal entity through which to conduct a business, that individual is considered to be operating the business as a sole proprietor, a status that does not give rise to a separate legal entity. Thus, the assets of the business are considered to be owned by the individual sole proprietor. In addition, that individual is liable for the debts and obligations of the business.¹¹⁶

¹¹² ULLCA § 303(a).

¹¹³ ULLCA §§ 301(a), (c).

¹¹⁴ See ULLCA § 303.

¹¹⁵ See, e.g., Burgess J.W. Raby & William L. Raby, *New Incentive for Avoiding SE and FICA Tax*, TAX NOTES TODAY, Dec. 10, 1998, 98 TNT 237-58.

¹¹⁶ For a discussion of the sole proprietorship, see Larry E. Ribstein, UNINCOR-

For purposes of the income tax, a sole proprietor is disregarded as a separate taxpaying entity.¹¹⁷ Thus, the business, as such, does not pay income tax on the profits generated by its activities. Instead, any profits from the business are included in the gross income of the owner.¹¹⁸ As a result, the owner pays tax on the business profits at the rates that apply to the owner.¹¹⁹

Aside from having to pay income tax on the business profits, a sole proprietor also will have to pay self-employment tax on those business profits.¹²⁰ Thus, as a general rule, any amounts that will be taxed to the owner for income tax purposes will also be taxed to the owner for employment tax purposes.

However, business losses are treated differently under the income and employment taxes. Under the income tax, the sole proprietor can utilize the loss to offset taxable income from other sources, both in the year of the loss and in other tax years.¹²¹ Thus, for example, if the business loses \$10,000 in its first year of operation, and if the owner has income from other sources in that year, the loss will offset \$10,000 of income from those other sources, triggering a reduction in the owner's income tax liability. However, if the owner has no income from other sources or if the income from those other sources is less than what he lost from the business, the owner can utilize that loss to offset taxable income in other years. The loss is generally utilized in the prior two tax years and the succeeding twenty tax years.¹²² A special rule permits

PORATED BUSINESS ENTITIES ch. 2.01 (3d ed. 2004).

¹¹⁷ Treas. Reg. § 301.7701-2(c)(2)(i). The mechanism for doing so is the Schedule C, which the owner must complete and include when filing her annual tax return.

¹¹⁸ I.R.C. § 61(a)(3).

¹¹⁹ Individuals are taxed under a graduated rate structure whose rates range from 10% to 35%. See *supra* text and table accompanying notes 30-33. The tax applies to amounts that exceed any exemptions, exclusions and deductions that are allowed to the individual. I.R.C. § 63.

¹²⁰ I.R.C. § 1402(a).

¹²¹ I.R.C. § 172(a).

¹²² I.R.C. § 172(b)(1)(A).

the individual to waive the use of a loss to offset income in prior years.¹²³ By permitting the owner to utilize a loss to offset income in years other than the one in which the loss arose, the income tax rules essentially represent an exception to the general requirement that individuals must compute their income tax liability on the basis of the activities and events that occur within specific annual accounting period.

By contrast, the employment tax rules generally do not permit a business loss to reduce the amounts subject to employment tax. Thus, if a sole proprietor has wages from employment in a year that he also sustained a \$10,000 loss from the business, the business loss will not offset any portion of the wages, which will be subject to the FICA tax to the extent they do not exceed the contribution and benefit base. However, if an individual has several businesses that are each operated as a sole proprietorship, a loss sustained by one business will be taken into account to determine the individual's net earnings from self-employment for that year.¹²⁴ However, if the loss from one business exceeds the profits from all the other businesses, the net loss still is not available to offset taxable amounts in other years. In this way, the employment tax rules treat multiple businesses as one enterprise, while continuing to base the individual's employment tax liability solely on the activities and events that occur during a particular annual accounting period.

C. Commentary on the Employment Tax Scheme

The existing federal employment tax statutory scheme is defective in several respects. First, it lacks a clear rule that applies to a member of any limited liability company that has not elected to be treated as a corporation for federal income tax purposes. The gap in the law gives rise to tax planning opportunities that can produce several undesirable outcomes. First, it increases the role that tax considerations play in the

¹²³ I.R.C. § 172(b)(3).

¹²⁴ Treas. Reg. § 1.1402(a)-2(c).

selection of an entity for conducting a business enterprise. Ideally, the choice of a business entity should not be influenced by tax considerations. Second, it creates the very real potential that individuals who may be in identical situations will interpret the law differently, leading to different computations of their employment tax liability. In an ideal world, individuals in substantially similar situations should pay the same tax. Third, the government is likely to collect less than what it is due in tax, leading to underfunding of the Social Security and Medicare programs at a time when their long term financial solvency is in question. The annual expenditures for the OASDI program will exceed employment tax income starting in 2017. At that time the social security trust fund will have to liquidate its assets to cover the shortfall until those assets are exhausted in 2040.¹²⁵ Fourth, the absence of clarity can increase how much it costs taxpayers to comply with the law. At one extreme, taxpayers who are willing and able to obtain competent tax advice will obtain such advice when they expect the tax savings to exceed any fees paid. Finally, the mere existence of a tax planning opportunity runs the risk of undermining respect for a tax system that relies on voluntary compliance, increasing the need for the government to fund enforcement activities.

Another defect in the existing statutory scheme is that the tax liability of a partner in a partnership depends on whether the partner owns a general or limited partnership interest. The use of two separate rules dates back to a time when there was a substantive difference between the two classes of partners. The self-employment tax rules initially required all partners (whether limited or general) to be taxed on their distributive share of income.¹²⁶ At the time, a limited partner had to be a passive investor in the partnership, or risk losing the protection of limited liability. Congress later determined that it was inappropriate to treat as earnings from work amounts

¹²⁵ Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds, Annual Report, at 2 (May 1, 2006).

¹²⁶ I.R.C. § 1402(a) (1949).

allocated to a limited partner who did not (and could not) play an active role in the business.¹²⁷ As a result, Congress changed the law so that the self-employment tax of a limited partner would only be based on amounts the partner received in exchange for rendering services to the business.¹²⁸ Over the course of time, however, state law has evolved to permit individuals to play active roles in an unincorporated business without losing the protection of limited liability.¹²⁹ As a result, it no longer makes sense for the self-employment tax to be based on the mere designation of a partner as a general or limited one.

Continued use of an outdated rule can cause individuals who are in otherwise similar situations to pay a different tax. Taxpayers can lose respect for the tax system when it produces such inconsistent outcomes. In addition, the continued reliance on an immaterial factor in the computation of tax liability can represent an undesirable tax planning opportunity that, because of its very existence, can further undermine the respect taxpayers have for the system. When taxpayers lose respect for the system, compliance rates are likely to suffer and income is likely to go underreported with greater frequency, making it necessary to commit more government funds to enforcement activities to reverse the trend.

A third defect is that different rules apply to determine the employment tax base of an employee-owner of a business depending on the business entity that is used to conduct the enterprise. Under current law, the SECA rules apply to sole proprietors and to any individual with an interest in a partnership or other unincorporated business entity that has not elected to be treated as a corporation for federal income tax purposes. Under those rules, the employment tax base can consist of the employee-owner's entire share of the profits of

¹²⁷ H.R. REP. NO. 95-702, pt. 1 (1977).

¹²⁸ I.R.C. § 1402(a)(13) (1949).

¹²⁹ These developments include adoption of legislation recognizing the limited liability company, the limited liability partnership, and the limited liability limited partnership.

the business, in addition to any compensation the individual receives for services rendered. By contrast, the FICA rules apply to any individual who owns an interest in any state law corporation or any unincorporated business entity that has elected to be treated as a corporation for federal income tax purposes. Under those rules, the tax base consists solely of amounts the employee-shareholder receives as compensation for services rendered. An individual's liability for employment tax should not depend on mere formalities.

Federal employment taxes are intended to apply to income from labor. When an individual works for a business he also owns and controls, it is difficult, if not impossible, to determine the extent to which the profits of the business are the result of the employee-owner's labor or a return on any capital that may be invested in the business. Wherever the line may fall, it seems unlikely to depend on the legal formalities that matter under the current statutory scheme. Moreover, whatever standard is used to draw the distinction, it should apply to all individuals who work for a business they also own and control. It should not matter whether the business operates through a corporation, partnership, limited liability company, sole proprietorship or some other vehicle yet to be created.

III. RECENT PROPOSAL FOR REFORM

A. Proposal of the Staff of the Joint Committee on Taxation

The Staff of the Joint Committee on Taxation (the JCT Staff) has advanced a proposal that goes a part of the way toward addressing these three defects.¹³⁰ The JCT Staff proposal would extend to any owner of a flow-through business entity the current set of rules that apply to general partners under the self-employment tax. Thus, all such owners would generally have to pay the self-employment tax on the profits of the business allocated to them, even if such income is not

¹³⁰ Joint Committee on Taxation, Options to Improve Tax Compliance and Reform Tax Expenditures 99-104 (Jan. 27, 2005) JCS-02-05.

actually paid out to them. These owners would also be taxable on any compensation received for services rendered to the business. However, if the owner does not materially participate in the business, only the owner's reasonable compensation from the entity is treated as subject to the SECA tax. This approach would apply to any owner in a business other than a sole proprietor or a shareholder in a taxable corporation.

The JCT Staff proposal retains the current limitation on the kinds of profits that are taxable to an owner under the self-employment tax. Thus, certain types of passive income like dividends and interest, certain gains, and other items that do not seem to qualify as income from labor would not be subject to tax. However, the proposal carves out an exception when the entity is in a service trade or business. In such situations, all of the profits allocated to an owner are treated as net earnings from self-employment.¹³¹

In his testimony at a May 5, 2005 Congressional hearing on Social Security, George K. Yin, Chief of Staff of the Joint Committee on Taxation explained the rationale for the rule.

The conceptual premise of the proposal is that the base for FICA and SECA taxes is labor income. The proposal applies this notion more uniformly than does present law to individuals who perform services for or on behalf of a pass-through entity in which they own an interest (i.e., a partnership, limited liability company, or S corporation). The proposal treats such individuals similarly to sole proprietors, as well as similarly to each other. Not only does this more uniform treatment improve the fairness of the tax law and increase the internal consistency of the tax rules, it also tends to improve tax neutrality by reducing the importance of FICA and SECA tax differences in taxpayers' choice of business entity.¹³²

¹³¹ A service business is defined to be one whose activities involve the performance of services in the fields of health, law, engineering, architecture, accounting, actuarial science, performing arts, or consulting. This definition is similar to the one that appears in I.R.C. § 448(d)(2).

¹³² *Hearing on "Social Security: Achieving Sustainable Solvency" Before the S.*

The JCT Staff proposal appears to adequately address the three principal defects in federal employment tax law in the following ways. First, the proposal eliminates the uncertainty over how a member of a limited liability company should determine his employment tax liability. Second, it disregards a partner's status as a general or limited partner as a factor in the computation of the partner's employment tax liability, substituting a rule that considers the degree to which the partner participates in the business of the partnership. Third, the proposal enunciates a uniform procedure for computing the employment tax liability of any individual who owns an interest in any nontaxable business entity. Thus, under the proposal, the provisions of the SECA would apply to determine the employment tax liability of anyone who owns an interest in an S corporation or any unincorporated entity that has not elected to be treated as a corporation for federal income tax purposes.

The JCT Staff proposal has been widely endorsed by the practicing bar for establishing a clear and uniform rule that applies to any employee-owner of a flow-through entity.¹³³ In addition, although there are many different views on how best to distinguish an employee-owner's labor income from any return on capital invested in a business, the organizations and individuals who have voiced their reaction to the proposal seem to believe that it strikes an appropriate balance between the interest in basing the tax on a precise measure of labor income, and the interest in using a rule that taxpayers can understand and apply with relative ease.¹³⁴

Comm. on Fin., 109th Cong. (2005) (statement of George K. Yin).

¹³³ See, e.g., New York State Bar Association Tax Section, Comments on JCT Recommendation Relating to Employment and Self-Employment Taxes of Partners, LLC Members and S Corporation Shareholders (Sept. 23, 2005) reprinted in Tax Notes (Oct. 31, 2005) at 675. However, the NYSBA believed the JCT Staff proposal defined the employment tax base too broadly. See also David C. Culpepper et al., *Self-Employment Taxes and Passthrough Entities: Where Are We Now?*, 91 TAX NOTES 211 (October 10, 2005) (endorsing the JCT Staff proposal with some reservations).

¹³⁴ See New York State Bar Association Tax Section, Comments on JCT Recommendation Relating to Employment and Self-Employment Taxes of Partners,

B. Commentary on the JCT Staff Proposal

The JCT Staff proposal does indeed represent an improvement over the current state of affairs. Among other things, by imposing the employment tax on amounts that a business makes for an employee-owner, not merely what the business pays to that person as compensation, the proposal severely reduces the opportunity for the understatement of labor income, and the underpayment of employment tax, by such individuals. However, because the rules do not apply to employee-shareholders of taxable corporations, the proposal falls short of establishing parity in the way federal employment tax rules apply to all individuals who conduct a business through a business entity.

The problems associated with this inconsistency are most apparent in the case of individual who plays a controlling role in the business that employs him. Such a business may generate profits that represent solely the product of the employee-owner's labor. However, under the JCT Staff proposal, if the business is conducted through a partnership, an employment tax would apply to the employee-partner's entire distributive share of the profits for the year, plus any salary. Meanwhile, if the business is conducted through a corporation, an employment tax would only apply to amounts actually paid to an employee-shareholder as compensation. That amount may be less than the individual's share of the business profits in any given year. In fact, it may be zero.

Indeed, in the setting of a corporation that is controlled by an employee-shareholder, that owner has an economic incentive to deal with the business in ways that minimize the total tax that both he and the corporation must pay. There is no reason this consideration would not play a role when the employee-shareholder wants to access the profits of the business and needs to decide how to do so. There could be many aspects to this decision, including whether the payout should take the

form of compensation or a dividend, the amount of the payout and when it should occur.

Before the temporary dividend tax cut, the ability of a controlling employee-shareholder to exploit this flexibility had limited practical significance for employment tax purposes. In almost all cases, the combined tax liability of the corporation and the individual would be kept to a minimum if a payout were structured as compensation.¹³⁵ Thus, there was almost no risk that a corporation would pay a dividend as a form of disguised compensation, potentially shortchanging Social Security and Medicare trust funds.¹³⁶ The primary advantage in the corporate setting was that a controlling employee-shareholder could decide whether to access the profits of the business at all. If he did not, none of the taxes associated with the transfer of money between him and the corporation would apply.

Ever since the temporary tax cut on dividends took effect, however, the math has changed. Now, if a controlling employee-shareholder wants to access the earnings of the corporation, that individual frequently has an incentive to do so by causing the corporation to pay him a dividend. To the extent the dividend is disguised compensation, the transaction avoids an employment tax that would otherwise apply to generate funds for social security and Medicare benefits at a time when the long term financial stability of those programs is at risk.

The following discussion illustrates the extent to which an incentive now exists for corporations to pay dividends as disguised compensation to controlling shareholders who work for the business. Among other things, the discussion shows that the dividend tax cut does not change the incentives in all cases involving closely held corporations. Rather, high income individuals who own and control low income corporations are the ones positioned to make the most of what appears to be an

¹³⁵ See *supra* text and table accompanying notes 38-39.

¹³⁶ Indeed, there was a greater risk that the corporation would overcompensate an employee-owner in order to secure a deduction that saved far more in income tax than it cost in employment tax.

unintended tax saving opportunity created by the temporary dividend tax cut.

IV. FINANCIAL INCENTIVES TO MINIMIZE FEDERAL EMPLOYMENT TAX

In order to illustrate the extent to which tax considerations can affect the form in which corporate profits are paid out to controlling employee-shareholders, this analysis considers the simplified case of a corporation that has only one shareholder. That individual also works for the company and desires to access \$15,000 of the corporation's earnings. He faces the choice of structuring the payout as a year-end bonus or as a dividend.¹³⁷ For purpose of the analysis, it is assumed the employee-shareholder is a married individual who files a joint tax return with his spouse. In addition, the analysis assumes that the business made no less than \$50,000 (before paying any compensation to the employee-owner) in the year of the payout.¹³⁸ Finally, the analysis assumes that any bonus paid by the corporation will be the employee-shareholder's only source of income subject to employment tax.

The analysis does not take into account any alternative minimum tax liability that may apply to the employee-owner or to the corporation.¹³⁹ In addition, although the phase out of deductions can affect the marginal rate that applies to the

¹³⁷ A shareholder can also receive a distribution in the form of a loan. Because a loan must be repaid, it is materially different from both a dividend and compensation. As a result, this analysis does not consider the tax consequences of a loan.

¹³⁸ As previously discussed, a corporation is taxed at 15% on taxable income up to \$50,000. See *supra* discussion accompanying notes 21-30. As that discussion points out, the marginal rate could go as high as 39% when taxable income falls between \$100,000 and \$335,000. By not assuming any ceiling on the corporation's earnings, the analysis leaves open the possibility that the corporation would fall anywhere within the full range of marginal tax rates that applies to corporations.

¹³⁹ A corporation subject to the alternative minimum tax would generally be taxed at a flat 20% on an adjusted taxable income figure referred to as alternative minimum taxable income. An individual subject to the alternative minimum tax is taxed under a two-tiered graduated rate structure with 26% and 28% as the rates. I.R.C. §§ 55-59 (2000).

income of an individual, such phase outs are not taken into account.¹⁴⁰

A. Compensation for Services

There are several tax effects produced by the payment of compensation by a corporation to an employee. First, the corporation can deduct amounts paid that are reasonable for the services rendered to it.¹⁴¹ Any amounts received by the employee-owner as compensation count as gross income to him, triggering an income tax liability.¹⁴² In addition, because the compensation qualifies as wages from employment, it also triggers an employment tax liability under FICA, with the employee and the corporation each being responsible for half the tax. The corporation is entitled to deduct its half of the tax paid.¹⁴³ That would reduce the income that is subject to the corporate tax, lowering the income tax liability of the business.¹⁴⁴ The following sections quantify the amount of tax owed or saved as a result of each of these effects.

1. Employment Tax Effects

The corporation will have to pay an amount equal to 6.2% of the bonus to cover its half of the OASDI component of the FICA tax. Thus, it will owe \$930 on a \$15,000 bonus payment.¹⁴⁵ Meanwhile, the employee would also have to pay \$930 to cover his portion of the tax on that bonus. Because the bonus is well below the contribution and benefit base, there is

¹⁴⁰ Robert J. Peroni, *Reform in the Use of Phase-Outs and Floors in the Individual Income Tax System*, 91 TAX NOTES 1415 (Special Supp., May 28, 2001).

¹⁴¹ I.R.C. § 162(a)(1).

¹⁴² I.R.C. § 61(a)(1).

¹⁴³ I.R.C. § 162(a).

¹⁴⁴ The amount of compensation paid would have ancillary consequences. If the corporation pays for health insurance for the employee and his family, the deduction available to the corporation could depend on the amount paid to the employee. In addition, the amount that the owner employee can receive as deferred compensation depends in part on the amount of compensation the owner employee receives. These ancillary consequences are not taken into account in the analysis.

¹⁴⁵ \$15,000 × 6.2% = \$930.

no possibility that any portion of the bonus would be exempt from the OASDI component of the tax.

The corporation and the employee-shareholder will each have to pay a 1.45% tax on the bonus to cover the hospital insurance component of FICA. Thus, a \$15,000 bonus will cost the company \$218 in tax, and it will also cost the employee-shareholder \$218 in tax.¹⁴⁶

2. Corporate Income Tax Effects

The corporation will be entitled to deduct from gross income any compensation it pays to its employee-owner.¹⁴⁷ In addition, the corporation will be entitled to deduct its share of any employment tax on that compensation.¹⁴⁸ These deductions will translate into a lower corporate income tax liability. The actual tax savings will depend on the tax that would otherwise be due on income that is offset by the deductions.

Because the corporate tax is imposed under a system of graduated marginal rates, the tax savings will depend on the tax bracket into which the corporation falls in the year it makes the payments. As previously discussed, there are six marginal rates, ranging from a low of 15% to a high of 39%. At the low end of the spectrum, if the corporation is in the 15% bracket, \$15,000 in business profits (unreduced by any bonus payment) would cost the corporation \$2,250 in income tax.¹⁴⁹ Conversely, if the corporation uses that money to pay a deductible bonus, there is no income left to be taxed, resulting in no income tax liability for the corporation on that money. Thus, a \$15,000 bonus payment would translate into \$2,250 in tax savings for a corporation in the 15% tax bracket. Meanwhile, at the high end of the spectrum, the same \$15,000 bonus would translate into \$5,850 of tax savings to a corporation in the 39%

¹⁴⁶ $\$15,000 \times 1.45\% = \$218$.

¹⁴⁷ I.R.C. § 162(a)(1). If the corporation were publicly traded, the deduction for salaries paid to certain executives would be limited to \$1 million. I.R.C. § 162(m).

¹⁴⁸ I.R.C. § 162(a).

¹⁴⁹ $\$15,000 \times 15\% = \$2,250$.

tax bracket.¹⁵⁰

The corporation would also be entitled to deduct any employment tax it must pay on any bonus paid to an employee. Like the deduction for the bonus itself, this deduction will also translate into tax savings that will vary with the corporation's marginal tax rate. There is a \$930 tax to cover the OASDI component of the FICA tax.¹⁵¹ That translates into \$140 in tax savings if the corporation is in the 15% marginal tax bracket.¹⁵² The savings top off at \$363 if the corporation is in the 39% marginal tax bracket.¹⁵³ For the health insurance component of the FICA tax, any \$15,000 of compensation would cost the corporation \$218 in tax that the corporation could deduct in computing its taxable income.¹⁵⁴ If the corporation is in the 15% tax bracket, that \$218 deduction corresponds to \$33 in income tax savings.¹⁵⁵ Meanwhile, if the corporation is in the 39% tax bracket, that \$218 deduction corresponds to \$85 in income tax savings.¹⁵⁶

3. Individual Income Tax Effects

Any bonus received by the employee-shareholder will be included in his gross income and subject to tax under the system of graduated marginal rates previously described.¹⁵⁷ For 2003 through 2010, the statute uses six marginal rates, ranging from a low of 10% to a high of 35%. After 2010, the pre-2001 marginal rate structure takes effect. It consisted of five marginal rates ranging from a low of 15% to a high of 39.6%. Thus, the income tax cost associated with the receipt of a bonus will depend on two things. First, it will depend on the year when the employee-shareholder receives the bonus. Second, it will depend on the employee-shareholder's marginal rate tax bracket in that

¹⁵⁰ $\$15,000 \times 39\% = \$5,850$.

¹⁵¹ See *supra* note 145.

¹⁵² $\$930 \times 15\% = \$140$.

¹⁵³ $\$930 \times 39\% = \$363$.

¹⁵⁴ See *supra* note 146.

¹⁵⁵ $\$218 \times 15\% = \$33$.

¹⁵⁶ $\$218 \times 39\% = \$85$.

¹⁵⁷ See *supra* discussion accompanying notes 26-32.

year. Assuming the \$15,000 bonus is received in 2006, the individual will have to pay as little as \$1,500 in tax if he is in the 10% tax bracket.¹⁵⁸ However, that \$15,000 bonus will cost him as much as \$5,250 in income tax if he is in the 35% tax bracket.¹⁵⁹

The following table summarizes the net savings and costs on a \$15,000 bonus paid by the corporation to the shareholder in 2003 through 2010. The net effect varies depending on two factors: the corporation's marginal tax rate and the shareholder's marginal tax rate.

**\$15,000 Bonus Paid by Corporation to Shareholder
Combined Tax Savings (Cost)
2003 through 2010**

Shareholder Marginal Rate						
Corp. Rate	10%	15%	25%	28%	33%	35%
15%	(\$1,373)	(\$2,123)	(\$3,623)	(\$4,073)	(\$4,823)	(\$5,123)
25%	\$242	(\$508)	(\$2,008)	(\$2,458)	(\$3,208)	(\$3,508)
34%	\$1,695	\$945	(\$555)	(\$1,005)	(\$1,755)	(\$2,055)
35%	\$1,857	\$1,107	(\$393)	(\$843)	(\$1,593)	(\$1,893)
38%	\$2,341	\$1,591	\$91	(\$359)	(\$1,109)	(\$1,409)
39%	\$2,503	\$1,753	\$253	(\$197)	(\$947)	(\$1,247)

Among other things, the table shows that there are situations in which the corporation and employee-shareholder collectively save more in taxes than they owe. The combined net savings is as high as \$2,503 when the corporation is in the 39% marginal tax bracket and the shareholder is in the 10% marginal tax bracket. As a practical matter, however, that particular pairing of tax brackets represents an anomalous situation.¹⁶⁰ Mean-

¹⁵⁸ $\$15,000 \times 10\% = \$1,500$.

¹⁵⁹ $\$15,000 \times 35\% = \$5,250$.

¹⁶⁰ The 10% tax bracket applies when an individual has taxable income that

while, in the vast majority of situations, the payment of compensation produces a net tax cost. The cost tops out at \$5,123 when the corporation is in the 15% marginal tax bracket and the shareholder is in the 35% marginal tax bracket.

After 2010, the range of outcomes will change because the existing schedule of marginal income tax rates that applies to individuals will be replaced with the schedule that had been in effect prior to 2001. That schedule used five rates ranging from 15% to 39.6%. This change will alter the calculus by increasing what the shareholder must pay in tax on the bonus. That will increase the combined net tax cost, or eliminate any combined net tax savings, that would have otherwise occurred under the existing schedule of marginal rates. The following table summarizes the range of outcomes produced after 2010 on the assumption that the pre-2001 schedule of marginal rates takes effect.

**\$15,000 Bonus Paid by Corporation to Shareholder
Combined Tax Savings (Cost)
After 2010**

Shareholder Marginal Rate

Corp. Rate	15%	28%	31%	36%	39.6%
15%	(\$2,123)	(\$4,073)	(\$4,523)	(\$5,273)	(\$5,813)
25%	(\$508)	(\$2,458)	(\$2,908)	(\$3,658)	(\$4,198)
34%	\$945	(\$1,005)	(\$1,455)	(\$2,205)	(\$2,745)
35%	\$1,107	(\$843)	(\$1,293)	(\$2,043)	(\$2,583)
38%	\$1,591	(\$359)	(\$809)	(\$1,559)	(\$2,099)
39%	\$1,753	(\$197)	(\$647)	(\$1,397)	(\$1,937)

The combined net tax effects associated with a bonus must

does not exceed \$15,100. That would mean that virtually all of the individual's other income was offset by exemptions, exclusions, and deductions of one kind or another. The 39% tax bracket applies when a corporation has taxable income over \$100,000 and up to \$335,000. I.R.C. § 11(b).

be compared to the combined net tax effects associated with a dividend. The following section computes the combined tax cost incurred by a corporation and its shareholder on the payment of a dividend.

B. Dividends

A payment of dividends triggers a different, and less complex, set of tax effects on to the corporation and its employee-shareholder. Understandably, there are no employment tax effects to consider. In addition, because the corporation is not entitled to deduct any dividends paid to shareholders, there are no corporate income tax effects to consider. The only income tax effects will occur at the level of the employee-shareholder.

Dividends received by the employee-owner will count as gross income to him.¹⁶¹ As such they will be subject to tax. As described above, the amount of tax will depend on two factors: the year of payment, and, where applicable, whether the dividend is a qualified dividend.¹⁶² This analysis assumes that any dividend received by the employee-owner will be a qualified dividend whenever applicable. For 2003 through 2007, qualified dividends are taxable at one of two rates, depending on the marginal tax rate that applies to the recipient's ordinary income. If the recipient is in the 25% bracket or higher, a qualified dividend is taxed at 15%.¹⁶³ If the recipient is any other tax bracket, a qualified dividend is taxed at 5%.¹⁶⁴ In 2008 through 2010, dividends are tax free to any recipient in a tax bracket below 25%, while the tax on dividends remains at 15% in all other cases.¹⁶⁵ The preferential rates that apply to qualified dividends are scheduled to expire after 2010.¹⁶⁶ Thus,

¹⁶¹ I.R.C. § 61(a)(7).

¹⁶² See *supra* discussion accompanying notes 25-39.

¹⁶³ I.R.C. § 1(h)(1)(C) (Supp. 2003).

¹⁶⁴ I.R.C. § 1(h)(1)(B) (Supp. 2003).

¹⁶⁵ *Id.*

¹⁶⁶ The Bush administration has proposed making permanent the preferential rates of tax that apply to qualified dividends. DEP'T OF THE TREASURY, GENERAL EXPLANATIONS OF THE ADMINISTRATION'S FISCAL YEAR 2007 REVENUE PROPOSALS 3 (February 2006).

the dividends would be subject to tax under the system of marginal rates that apply to all ordinary income.

This array of transitory rules makes it necessary to compute the combined net tax effects for many different periods of time. A \$15,000 dividend paid in 2003 through 2007 will have a net tax cost of \$2,250 if the recipient is in the 25% or higher marginal tax bracket. The net tax cost will be \$750 if the recipient is below the 25% tax bracket. The following table summarizes the full range of outcomes.

**\$15,000 Dividend Paid by Corporation to Shareholder
Combined Tax Savings (Cost)
2003 through 2007**

Shareholder Marginal Rate

Corp. Rate	10%	15%	25%	28%	33%	35%
All Rates	(\$750)	(\$750)	(\$2,250)	(\$2,250)	(\$2,250)	(\$2,250)

A \$15,000 dividend paid in 2008 through 2010 will continue to have a net tax cost of \$2,250 if the recipient is in the 25% or higher marginal tax bracket. However, there will be no tax cost if the recipient is below the 25% tax bracket. The following table summarizes the full range of outcomes.

**\$15,000 Dividend Paid by Corporation to Shareholder
Combined Tax Savings (Cost)
2008 through 2010**

Shareholder Marginal Rate

Corp. Rate	10%	15%	25%	28%	33%	35%
All Rates	\$0	\$0	(\$2,250)	(\$2,250)	(\$2,250)	(\$2,250)

A \$15,000 dividend paid after 2010 will continue to have a net tax cost that depends on the recipient's marginal tax bracket.

et. However, because the pre-2001 tax brackets will replace the ones now in effect, the range of outcomes will be different. At the low end of the spectrum, there will be a net tax cost of \$2,250 if the recipient is in the 15% tax bracket. The net tax cost will be as high as \$5,940 if the recipient is in the 39.6% tax bracket. The following chart summarizes the full range of outcomes.

**\$15,000 Dividend Paid by Corporation to Shareholder
Combined Tax Savings (Cost)
After 2010**

Shareholder Marginal Rate					
Corp. Rate	15%	28%	31%	36%	39.6%
All Rates	(\$2,250)	(\$4,200)	(\$4,650)	(\$5,400)	(\$5,940)

C. Quantifying the Incentive

The difference between the combined net tax effects associated with each payout alternative will determine the extent to which there is a tax incentive for structuring a payout as one or the other. Because the combined net tax effects produced in any case varies depending on a number of transitory rules, any comparison between the two payout options is valid for only a discrete period of time. This analysis proceeds accordingly, starting with an analysis of the period covered by 2003 through 2007.

In 2003 through 2007, qualified dividends are taxed as long term capital gains, and ordinary income is subject to tax under a temporary schedule of marginal rates ranging from 10% to 35%. The following table shows the extent to which the corporation and the shareholder collectively save more, or pay more, in tax when a \$15,000 payout is structured as a bonus instead of a dividend.

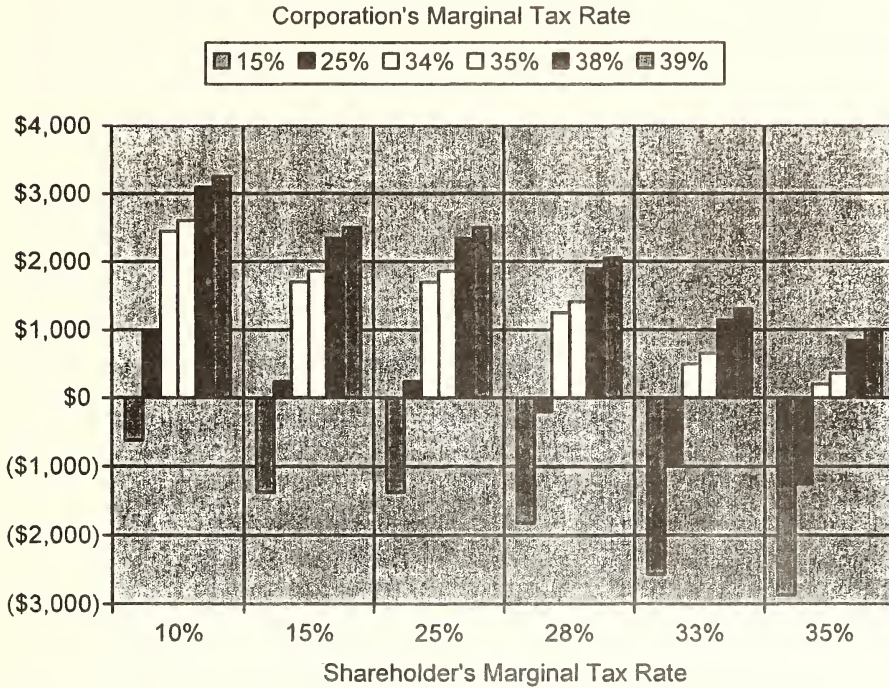
**Tax Advantage of Paying \$15,000
As a Bonus Instead of Dividends
2003 through 2007**

Shareholder Marginal Rate

Corp. Rate	10%	15%	25%	28%	33%	35%
15%	(\$623)	(\$1,373)	(\$1,373)	(\$1,823)	(\$2,573)	(\$2,873)
25%	\$992	\$242	\$242	(\$208)	(\$958)	(\$1,258)
34%	\$2,445	\$1,695	\$1,695	\$1,245	\$495	\$195
35%	\$2,607	\$1,857	\$1,857	\$1,407	\$657	\$357
38%	\$3,091	\$2,341	\$2,341	\$1,891	\$1,141	\$841
39%	\$3,253	\$2,503	\$2,503	\$2,053	\$1,303	\$1,003

The table indicates that a dividend costs more in tax than does a bonus in the vast majority of situations. For instance, a \$15,000 bonus enjoys a \$1,003 tax advantage over a \$15,000 dividend when the shareholder is in the 35% marginal tax bracket and the corporation is in the 39% marginal tax bracket. In other words, in that situation the corporation and shareholder end up with \$1,003 more after tax by structuring the payout as a bonus as opposed to a dividend. However, there is a range of situations when a dividend enjoys a tax advantage over a bonus. Specifically, that advantage exists whenever the corporation is in the 15% marginal tax bracket. It also exists when the corporation is in the 25% tax bracket and the shareholder is in one of the three highest income brackets. In those situations, there is a financial incentive to substitute a dividend for a bonus in order to minimize the net tax cost to the shareholder and corporation. The following chart depicts the range of outcomes graphically.

Figure 1
Tax Advantage to Paying \$15,000
As a Bonus instead of Dividends
2003 through 2007



The bars that rise above the zero dollar line represent the cases when a \$15,000 bonus costs less tax than a dividend of the same amount. Conversely, the bars that fall below the zero dollar line represent cases when a \$15,000 dividend costs less than a bonus of the same amount. Thus, a shareholder-employee in the 35% tax bracket will pay less in tax on a bonus, compared to a dividend, when the corporation is in any tax bracket above 25%. However, that same shareholder-employee will pay less in tax on a dividend, compared to a bonus, when the corporation is in any tax bracket below 34%. The length of the bar corresponds to the amount of tax dollars saved in any given case. Thus, if the shareholder is in the 35% tax bracket and the

corporation is in the 15% tax bracket, a dividend will cost nearly \$3,000 less in tax compared to a bonus.

The graph makes it easier to appreciate the distribution of benefits that are available when a dividend is substituted for a bonus. Individuals in the three highest income brackets can save tax dollars by making the substitution in a wider range of situations. In addition, when they do so, they save far more in taxes than their lower income counterparts. In effect, the rich can get richer when they work for free for a corporation they own and control.

The range of outcomes changes slightly in 2008 through 2010 when qualified dividends are tax exempt to any recipient whose ordinary income is taxed below 25%. Because the change in the law reduces the cost associated with making a dividend, the incentive to substitute a dividend for a bonus operates in a greater number of cases affected by the tax cut. The following table quantifies the economic incentive to make the dividend-for-bonus substitution in 2008 through 2010.

**Tax Advantage of Paying \$15,000
As a Bonus instead of Dividends
2008 through 2010**

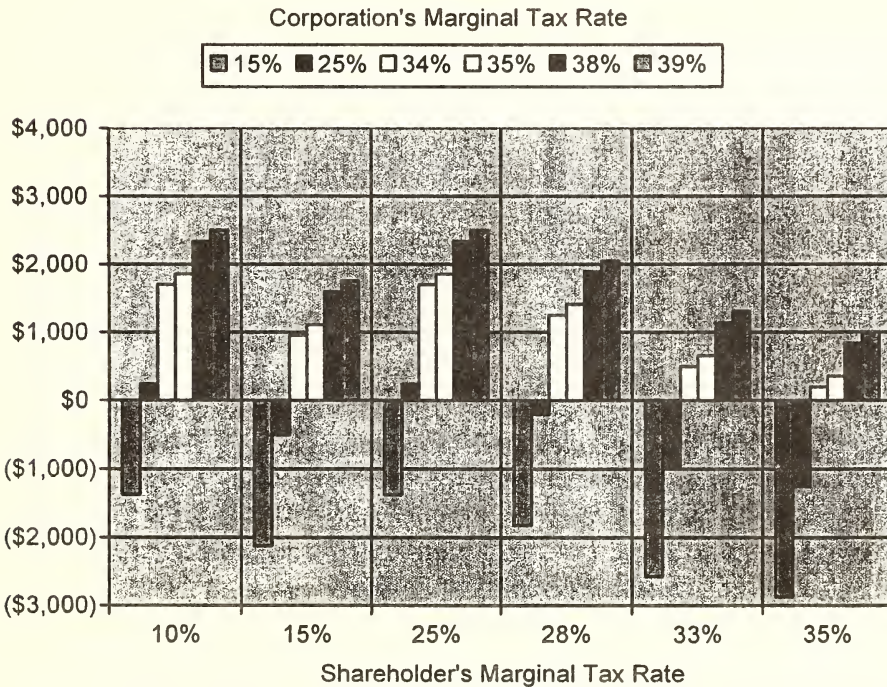
Shareholder Marginal Rate

Corp. Rate	10%	15%	25%	28%	33%	35%
15%	(\$1,373)	(\$2,123)	(\$1,373)	(\$1,823)	(\$2,573)	(\$2,873)
25%	\$242	(\$508)	\$242	(\$208)	(\$958)	(\$1,258)
34%	\$1,695	\$945	\$1,695	\$1,245	\$495	\$195
35%	\$1,857	\$1,107	\$1,857	\$1,407	\$657	\$357
38%	\$2,341	\$1,591	\$2,341	\$1,891	\$1,141	\$841
39%	\$2,503	\$1,753	\$2,503	\$2,053	\$1,303	\$1,003

The chart shows that if the employee-shareholder is in the 15% tax bracket, there will be an incentive to substitute a dividend for a bonus in cases where the corporation is in the 25% tax bracket. The following chart depicts the full range of out-

comes graphically, permitting one to more easily appreciate the distribution of benefits available under a dividend-for-bonus substitution.

Figure 2
Tax Advantage to Paying \$15,000
In Compensation instead of Dividends
2008 through 2010



Prior to 2008, the dividend-for-bonus substitution produced greater tax savings more frequently as the employee-shareholder moved into higher income tax brackets. The temporary divi-

dend tax exemption that applies in 2008 through 2010 makes it possible for some lower income individuals to save more tax dollars in a greater number of cases than they would have otherwise. However, individuals in the highest income brackets continue to enjoy the lion's share of any tax savings by working for free.

If the dividend tax cut expires after 2010, the incentive to substitute dividends for a bonus disappears. The following table summarizes the extent to which a \$15,000 bonus has a tax advantage over a \$15,000 dividend after 2010.

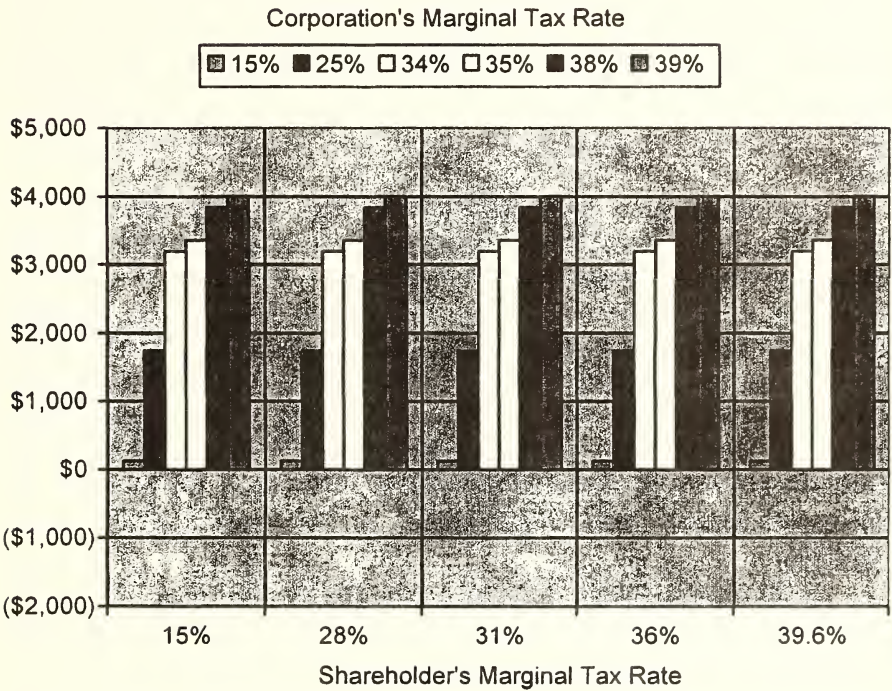
**Tax Advantage of Paying \$15,000
As a Bonus Instead of Dividends
After 2010**

Shareholder Marginal Rate

Corp. Rate	10%	15%	25%	28%	33%	35%
15%	\$127	\$127	\$127	\$127	\$127	\$127
25%	\$1,742	\$1,742	\$1,742	\$1,742	\$1,742	\$1,742
34%	\$3,195	\$3,195	\$3,195	\$3,195	\$3,195	\$3,195
35%	\$3,357	\$3,357	\$3,357	\$3,357	\$3,357	\$3,357
38%	\$3,841	\$3,841	\$3,841	\$3,841	\$3,841	\$3,841
39%	\$4,003	\$4,003	\$4,003	\$4,003	\$4,003	\$4,003

The following chart depicts these outcomes graphically.

Figure 3
Tax Advantage to Paying \$15,000
In a Bonus instead of Dividends
After 2010



The chart underscores the fact that once the dividend tax cuts expire, there is no tax related incentive to pay a dividend as disguised compensation. Therefore, the tradeoffs that had been in effect before the tax cuts were introduced will apply again.

D. The Unequal Opportunity Tax Loophole

Clearly any tax savings to be realized by substituting dividends for a bonus will be enjoyed primarily by high income individuals. As the charts and tables show, the most tax savings occur when the employee-shareholder is in the highest tax brackets and the corporation is in the two lowest tax brackets. There is no concrete evidence that shows the extent to which high income individuals own and control corporations that make low incomes. However, there is compelling evidence that ownership of closely held corporations is severely concentrated in the hands of the most wealthy individuals. It also seems clear that all but a small minority of corporations have incomes low enough to place them in one of the two lowest tax brackets.¹⁶⁷

The available evidence shows that the stock in closely held corporations is concentrated in the hands of very wealthy individuals. The Internal Revenue Service estimates that there were 7.4 million individuals in the United States with at least \$675,000 in gross assets in 2001, representing 3.5% of the total U.S. adult population.¹⁶⁸ Nearly 1.2 million of this group of wealthy individuals, representing over 16% of the total, owned stock in non-publicly traded corporations.¹⁶⁹ The total value of this stock was estimated to be over \$1.2 trillion. However, such stock ownership was concentrated in the hands of the very wealthy.

Individuals whose personal net worth exceeded \$2.5 million accounted for 12.8% of wealthy individuals, but they owned

¹⁶⁷ The ownership and income of corporations was previously analyzed by John W. Lee, *A Populist Political Perspective of the Business Tax Entities Universe: "Hey the Stars Might Lie But the Numbers Never Do,"* 78 TEX. L. REV. 885 (2000).

¹⁶⁸ Barry W. Johnson & Brian G. Raub, *Personal Wealth, 2001*, STATISTICS OF INCOME BULLETIN, Winter 2005–2006, at 120. These estimates are based on the gross assets reported in federal estate tax returns, the only source of data from which to estimate the wealth holdings of the general population. *Id.* The term "gross assets reflects the gross value of all assets owned by a decedent, including the full face value of life insurance, reduced by the value of any policy loans." *Id.* The figure is not adjusted to reflect any other debts owed by the decedent. *Id.*

¹⁶⁹ *Id.* at 136, tbl. 1.

76.8% of the value of all stock in non-publicly traded corporations owned by wealthy individuals. Individuals whose personal net worth exceeded \$5 million accounted for 5.0% of all wealthy individuals, but they owned 63.3% of all stock in non-publicly traded corporations.¹⁷⁰ The following table details the distribution of non-publicly traded stock among individuals whose gross assets exceed \$625,000.¹⁷¹

**Ownership of Non-Publicly Traded Stock
by Owner's Size of Net Worth
2001**

Owner's of Net Worth	Number in Class	% to Total	Total Value of Stock	% to
	(000)		(000,000)	
\$20,000,000 or more	46	0.6%	449,612	36.6%
\$10,000,000 under \$20,000,000	77	1.0%	154,327	12.6%
\$5,000,000 under \$10,000,000	243	3.3%	173,792	14.1%

¹⁷⁰ This pattern of ownership is consistent with the findings of other studies. The Federal Reserve Board conducts a survey of consumer finances every three years. The 2004 survey showed that ownership of privately held business increases with family income. Brian K. Bucks, Arthur B. Knuckle & Kevin B. Moore, *Recent Changes in U.S. Family Finances: Evidence from the 2001 and 2004 Survey of Consumer Finances*, FEDERAL RESERVE BULLETIN-2006, March 22, 2006, at A22, tbl. 8B. Less than 4% of the families in the bottom quintile owned equity in a business, while over a third of the families in the top tenth owned equity in a business. *Id.* The survey considers equity in a business to include sole proprietorships, limited partnerships, other types of partnerships, subchapter S corporations and other types of corporations that are not publicly traded, limited liability companies and other types of private businesses. *Id.* at A24, n. 30. However, the category does not include self-employed individuals. *Id.*

¹⁷¹ Johnson and Raub, *supra* note 168, at 136, tbl. 1. An individual's net worth is equal to the value of the person's gross assets reduced by any debts that the individual owes. *Id.* at 120.

\$2,500,000 under \$5,000,000	574	7.8%	165,506	13.5%
\$1,000,000 under \$2,500,000	2,569	34.9%	203,715	16.6%
\$600,000 under \$1,000,000	2,307	31.4%	49,671	4.0%
Under \$600,000 under \$1,000,000	1,509	20.5%	29,826	2.4%

Not only is the distribution of closely held corporations concentrated in the hands of wealthy individuals, but also there is strong evidence showing that the vast majority corporations have low incomes.

Statistics compiled by the Internal Revenue Service indicate that low income corporations (other than S corporations) account for the overwhelming share of all active corporations.¹⁷² The agency's annual survey of corporate incomes classifies corporations by the size of their assets.¹⁷³ By this measure, non-S corporations are concentrated at the low end of the spectrum.¹⁷⁴ According to data from the 2003 survey, nearly 75% of all active corporations in this group have less

¹⁷² An active corporation is one that reported either an item of income or an item of deduction or both on the federal income tax return. Internal Revenue Service, *Section 5: Explanation of Terms*, 2002 CORPORATION COMPLETE REPORT, at 261.

¹⁷³ The term "total assets" refers to amounts reported by a corporation in the end-of-year balance sheet. The figure is a net amount that reflects reductions for accumulated depreciation, accumulated amortization, accumulated depletion, and the company's reserve for bad debts. *Id.* at 272.

¹⁷⁴ This universe of corporations includes the conventional domestic C corporation. It also includes certain life insurance companies, property and casualty insurance companies, regulated investment companies (mutual funds), real estate investment trusts, and foreign corporations with business income from U.S. sources. Internal Revenue Service, *Section 3: Description of the Sample and Limitations of the Data*, CORPORATION COMPLETE REPORT (2002) at 7.

than \$500,000 in assets, nearly 85% have under \$1 million in assets, and nearly 95% have under \$5 million in assets.

Moreover, these low asset corporations that dominate the universe also seem to have low incomes. Over two-thirds of active non-S corporations have no more than \$500,000 in assets. The average net income of that group of non-S corporations is just over \$20,000. Roughly 10% of corporations have at least \$500,000 in assets but not more than \$1 million in assets. The average net income of that group is just over \$54,000, which is slightly above the upper limit of the lowest tax bracket: 15%.¹⁷⁵ The following table shows the average net income for all corporations in each asset range for 2003.¹⁷⁶ Among other things, the table shows that low asset corporations dominate the scene and that there is a strong correlation between the size of a corporation's assets and the net income it generates.

¹⁷⁵ It would be much more useful to know the number of corporations reporting income not in excess of each of the tax rate thresholds specified by statute. However, the data is not collected or reported in a way that permits that kind of analysis. Nevertheless, this less-than-perfect tabulation of the available data is sufficient to make a convincing case that low income corporations account for the lion's share of the total universe of corporations. This conclusion is consistent with observations made by others. *See, e.g., Lee, supra* note 167, at 906.

¹⁷⁶ Internal Revenue Service, 2003 CORPORATION SOURCE BOOK OF STATISTICS OF INCOME, Publication 1053, May 2006. The publication includes a master table reflecting statistics for all corporations with net income, including S corporations. *Id.* at tbl. 2. That table also shows the portion of net income represented by S corporations. A separate table provides statistics solely for S corporations with net income. *Id.* at tbl. 4. It shows, among other things, the number of S corporations falling into that category. *Id.*

**Active Non-S Corporations with Net Income
2003**

Asset Size	Number	% to Total	Net Income	Average Net Income
Zero Assets	75,404	7.5%	14,166,113	187,870
\$1 under \$500,000	672,271	67.2%	14,002,197	20,828
\$500,000 under \$1,000,000	95,246	9.5%	5,168,245	54,262
\$1,000,000 under \$5,000,000	105,471	10.5%	14,450,090	137,005
\$5,000,000 under \$10,000,000	16,900	1.7%	8,024,481	474,821
\$10,000,000 under \$25,000,000	11,291	1.1%	11,882,878	1,052,420
\$25,000,000 under \$50,000,000	5,340	0.5%	11,278,283	2,112,038
\$50,000,000 under \$100,000,000	4,503	0.5%	15,012,531	3,333,895
\$100,000,000 or more	13,688	1.4%	844,949,606	61,729,223

There is good reason to expect that controlling owner-employees of C corporations will exploit the opportunity to substitute dividends for a salary when doing so will save tax dollars.¹⁷⁷ Consider how S corporations have been used by their

¹⁷⁷ There is evidence that dividend income increased following the dividend tax cut. Qualified dividend income rose by 29.2% and the number of returns reporting that item increased by 9.2% in 2004, the first full year that the dividend tax cuts were in effect. Internal Revenue Service, *Section 1: Introduction and Changes in*

owners. In the year 2000, owners of 36,000 single-shareholder S corporations received no salaries at all even though the corporation in each case had over \$100,000 in operating profits.¹⁷⁸ Moreover, the owners of single-shareholder S corporations have been setting their salaries at a decreasing percentage of corporate profits. In 1994, these shareholders received salaries equal to 47.1% of the corporation's profits, but in 2001, the salaries fell to 41.5% of the profits.¹⁷⁹ The situation has been described in a Treasury Department Audit Report as a "multibillion dollar employment tax shelter."¹⁸⁰ The C corporation form of ownership will likewise operate as an employment tax shelter to many high income individuals as long as tax dollars can be saved by substituting dividends for a salary.

It is not unusual to find the owners of single shareholder S corporations working for free (or nearly so) because taxpayers and their tax advisors know it is extremely unlikely that the government will detect and punish the practice. In 2004, the Internal Revenue Service examined a meager .19% of S corporation returns, down from an equally meager .43% in 2001.¹⁸¹ Small corporations have been examined at a similarly low rate. The IRS examined .32% of small corporations in 2004, compared to .60% in 2001.¹⁸² To make matters worse, even when

Law, INDIVIDUAL INCOME TAX RETURNS PUBLICATION 1304, 2004, at 6-7, tbl. A. However, it is difficult to establish the extent to which any dividend income is a substitute for wages that a business would have otherwise paid an owner employee.

¹⁷⁸ TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION, *Actions Are Needed to Eliminate Inequities in the Employment Tax Liabilities of Sole Proprietorships and Single-Shareholder S Corporations*, Ref. No. 2005-30-080 (May 2005) at 12.

¹⁷⁹ *Id.* at 5.

¹⁸⁰ *Id.* It reports that single owner S corporations would have owed \$5.7 billion more in employment taxes in 2000 had those businesses operated as sole proprietorships, where all of the earnings would be subject to SECA. *Id.* at 6.

¹⁸¹ TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION, *The Small Business/Self-Employed Division is Beginning to Address Challenges that Affect Corporate Return Examination Coverage*, Ref. No. 2005-30-130, (August 2005) at 1. In a recent reported case, the court recast dividends paid by the corporation to its sole shareholder as wages subject to FICA taxes. *Nu-Look Design, Inc. v. Comm'r*, 356 F.3d 290 (3d Cir. 2004).

¹⁸² TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION, *supra* note 178, at

an S corporation is selected for audit, the agent often fails to examine the issue of officer compensation. A 2002 study revealed that agents failed to consider the adequacy of officer compensation in 22% of audited S corporations that made a distribution to their shareholders.¹⁸³ Equally alarming is the fact that the corporations examined in the study paid their officers an average of \$5,300 in wages compared to an average of \$340,323 in nonwage distributions, strongly suggesting a practice of substituting nonwage distributions for salaries in order to avoid employment tax. There is already a cottage industry built around advising small businesses to save on employment taxes by forming S corporations.¹⁸⁴ That cottage industry may only expand in scope now that the C corporation can be used to produce the same results with equally low rates of detection and punishment.¹⁸⁵

Thus, the opportunity to save taxes by substituting dividends for compensation is not an abstract matter. Working for free is a very real option confronted by what appears to be a substantial number of individuals who own and control a corporation that employs them. Moreover, there is compelling evidence that the opportunity to save taxes is not shared uniformly within the universe of owners of closely held businesses. The rich stand more to gain, and there appears to be very few other individuals who are in a position to exploit this opportunity. To

1. Any corporation with less than \$10 million in assets is classified as a small corporation. *Id.*

¹⁸³ TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION, *The Internal Revenue Service Does Not Always Address Subchapter S Corporation Officer Compensation During Examinations*, Ref. No. 2002-30-125 (July 2002) at 3-4.

¹⁸⁴ TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION, *supra* note 178, at 13. There is no shortage of literature discussing ways to capitalize on this opportunity to minimize or eliminate employment taxes where it would otherwise apply. See, e.g., James L. Wittenbach and Ken Milani, *FICA Factors for S Corporation Payments to Owner/Employees*, 75 PRAC. TAX STRATEGIES 338 (December 2005).

¹⁸⁵ In fact there is already practitioner oriented literature discussing the situations in which the owner of a closely held corporation can save tax dollars by substituting dividends for a salary. See, e.g., Thomas Zupanc and Sabyasachi Basu, *Re-Evaluate the Various Ways to Tap Corporate Cash*, 74 PRAC. TAX STRATEGIES 332 (June 2005).

compound the situation, there seems to be a very low risk that the government will detect and punish instances of people working for free. In a very real sense, the dividend tax cut has inadvertently created a rich man's tax loophole that reduces employment tax receipts at precisely the time when the social security and Medicare trust funds need the money more than ever.

V. EXPOSING AND REFORMING EMPLOYMENT TAX DEFECTS

There is no economic incentive to substitute a dividend for a bonus after the dividend tax cut expires. However, even in that situation, the corporate form offers employment tax advantages that other forms of doing business do not. Most important, a corporation can control when (if ever) the employment tax is triggered. There is no employment tax liability as long as the corporation does not pay compensation to an employee-owner. In the event compensation is paid, the employment tax liability will be based on the amount paid. Moreover, because compensation in excess of the FICA contribution and benefit base is exempt from the OASDI component of the tax, employment taxes could be saved by compressing multiple years worth of compensation into a single year. Thus, if the owner received \$180,000 in compensation in 2006, only \$94,200 would be subject to the 12.4% OASDI tax. The rest would be exempt from that tax, even though it may relate to services performed during a year when the corporation did not pay the owner a salary. Therefore, even when the employment tax is triggered, the tax liability can be managed and minimized by an individual who owns and controls the corporation that employs him.

The JCT Staff proposal comes close to eliminating any opportunity for understating employment tax liability when an individual owns and works for a business conducted through a flow-through entity. That individual is subject to employment tax on any amounts paid as compensation and on the individual's share of the profits of the business, provided that

he materially participates in the business.¹⁸⁶ However, the JCT Staff proposal leaves in place the opportunities to understate employment tax liability when an individual owns and works for a business conducted through a corporation. Such an individual is only subject to employment tax on amounts actually paid as compensation. Allowing this rule to remain in effect ignores the full range of opportunities to control and minimize employment tax liability when a self-employed individual operates through a corporation.

It makes little sense for an individual's employment tax liability to depend on the business form through which the business is conducted. However, that is exactly what happens now, and it will continue to happen, with less frequency, in the event the JCT Staff proposal is adopted. Formal distinctions ought to matter least in the case of a closely held business, where the interests of the business and a controlling employee-owner are not adverse. Transactions between the two parties are unlikely to occur at arms length. The JCT Staff proposal prevents the employment tax from being a victim of self-dealing when a flow-through entity is used to conduct a business. However, the employment tax falls victim to self-dealing when an individual operates his business through a corporation. Not only does this represent an undesirable tax planning opportunity, it causes taxpayers who are in economically similar situations to be treated differently solely because the rules operate by reference to irrelevant legal formalities.

Federal employment tax rules would produce more sensible and uniform outcomes if economic realities, not legal formalities, played a greater role in determining how the rules apply. There are real and compelling ways for an individual to understate his employment tax liability when he owns a corporation that employs him. These opportunities are more likely to be

¹⁸⁶ Determining whether someone materially participates in a business depends on a wide range of factors. However, the regulations apply several bright line standards, making it difficult, but not impossible, for someone to understate his employment tax liability by misrepresenting the role he plays in the business. See *Treas. Reg. § 1.469-5T* (2006).

exploited when the individual is in a position to influence decisions by the corporation. Thus, any approach to address this situation should consider the extent to which such opportunities to exercise and exploit control exist. As a general proposition, such opportunities occur in the closely held corporation.¹⁸⁷

The JCT Staff proposal could be improved if its rules were made to apply to individuals who own and work for closely held corporations. There are several ways to define a closely held corporation. However, there are already a number of instances in which a corporation is considered to be closely held for income tax purposes if the corporation satisfies the stock ownership test contained in the personal holding company rules.¹⁸⁸ Under that test, a corporation would be considered closely held if five or fewer individuals own more than 50% of the stock of the corporation during the last six months of the taxable year.¹⁸⁹

Whenever a corporation qualifies as a closely held corporation, then the rules of SECA should apply to determine the employment tax liability of any owner, as follows. First, the tax should apply to the investor's share of the corporation's taxable income for any given year and to any amounts paid to the owner as compensation for services rendered. A corporation already computes its taxable income each year; the employee-owner's share would simply be a function of the shareholder's interest in the corporation as measured by stock owned. Thus, if the employee-shareholder owned 40% of the stock in a year that the corporation generated \$160,000 in earnings and profits, \$64,000 of that amount would represent that individual's share of the corporation's earnings.¹⁹⁰ In addition, the self-employment tax would apply to any amounts actually paid to that

¹⁸⁷ See Lee, *supra* note 167 (describing how closely held corporations are often used to secure tax advantages that are not available to other business entities).

¹⁸⁸ The personal holding company stock ownership test is incorporated in the at risk rules. I.R.C. § 465. The test is also incorporated in the passive activity loss limitation rules. I.R.C. § 469A.

¹⁸⁹ See I.R.C. § 542(a)(2).

¹⁹⁰ $\$160,000 \times 40\% = \$64,000$.

individual as compensation. However, if the shareholder does not materially participate in the business, only amounts actually paid to him as reasonable compensation would be subject to the SECA tax. This approach would establish near complete parity in the way the rules operate, regardless of the legal entity through which an individual conducts a business.

The JCT Staff proposal applies two sets of rules to determine the extent to which the profits of the business are subject to employment tax. For all businesses other than a service business, the employment tax applies only to the investor's share of earnings other than certain passive income items that are already excluded from the definition of net earnings from self-employment. However, if the business is in a service business, no such adjustment is made.¹⁹¹ The distinction between service and non-service businesses business should apply with equal force when the business is conducted through a corporation.

The suggestions just proposed offers a number of advantages. First, they would reduce the opportunities for undesirable tax planning that can shortchange the federal government in one way or another. Second, they would promote greater consistency in the way taxpayers are treated, resulting in more sensible outcomes. Finally, they would extend an element of economic reality to the way closely held businesses are addressed by federal tax law.

There are a number of instances in which special rules already apply to prevent taxpayers from exploiting opportunities to utilize corporations to avoid or evade tax. This is accomplished in a number of ways. In some cases the approach is not to treat the corporation as a separate and distinct taxpaying unit. In other cases, the approach is simply to penalize or eliminate any advantage that may be gained by using a corporation

¹⁹¹ The JCT Staff proposal employs an existing definition of a service business. Under that definition, a service business is one in which substantially all of the activities involve the performance of services in the fields of health, law, engineering, architecture, accounting, actuarial science, performing arts or consulting. I.R.C. § 448(d)(2)(A).

as an element in a particular arrangement. In all instances, however, there is an element of economic reality that informs the approach. To that extent, this proposal is merely a variation of existing approaches directed at combating taxpayer abuse of the corporate form. The following sections make this point clear.

A. Use of Multiple Corporations

As a general rule, a corporation is considered to be a separate tax paying unit, and higher tax rates are more likely to apply to a corporation's income as it earns more. However the income that would otherwise be reported by one corporation could be fractured if multiple corporations were used to generate that same income. If structured properly, such an arrangement could prevent the higher tax rates from coming into play, provided the amount of income generated by each commonly controlled corporation did not exceed the \$50,000 threshold that applies to the 15% marginal tax bracket. To combat against this maneuver, federal tax law will not recognize the independent tax status of a corporation if it and one or more others are under common control. Instead, all corporations under common control are treated as one for purposes of computing the income tax on their total income.¹⁹²

In operation, the rule disregards legal distinctions and instead uses economic realities as the basis for identifying the proper taxpaying unit. In a similar fashion, the proposal advanced above acknowledges the need to disregard the separate existence of the corporation when it is owned and controlled by someone who works for it. In addition, the proposal takes an approach that prevents someone from using a corporation to minimize or eliminate a tax liability that would otherwise apply.

¹⁹² I.R.C. § 1561(a).

B. Personal Services Corporation

As a general rule, a corporation and its owner are two separate and distinct taxpaying units. However, if the principal activity of the corporation is to perform personal services that are actually performed by a 10% owner, the Internal Revenue Service can reallocate items between the corporation and its owner, provided the arrangement was used to avoid tax.¹⁹³ Any allocation made by the government must be necessary to prevent avoidance or evasion of federal income tax or to clearly reflect the income of the corporation or its employee-owner.¹⁹⁴ The statute effectively acknowledges that transactions between a corporation and someone who controls it may not occur on an arm's length basis. To address this possibility, the statute enables the Internal Revenue Service to disregard the actual form of a transaction and to make reallocations that reflect the economic realities.

The proposal advanced in this article operates in a similar way. It manifestly acknowledges that transactions between corporations and the owners who control them should not be taken at face value. Thus, a payment that takes the form of a dividend should not be treated as such if it in fact constitutes disguised compensation for services rendered. Moreover, even if a payment is made in the form of compensation, the amount paid may bear little or no resemblance to the economic realities. Those economic realities may be difficult to determine in any situation where an individual is both an owner and an employee of a business. Because the earnings of the business

¹⁹³ I.R.C. § 269A(b). This rule does not authorize the Secretary to make a reallocation in the case under review. First, because the statute limits the Secretary's reallocation powers to cases involving the possible avoidance or evasion of federal income tax, the Secretary has no power to make a reallocation where the employment tax may be avoided or evaded. Second, even if the power were available to address possible evasion or avoidance of the employment tax, the power could only be exercised when the corporation performs services for one other corporation, partnership or other entity. As a result, the power cannot be applied to address the employee-owner of a corporation that performs services for more than one corporation, partnership or other entity.

¹⁹⁴ I.R.C. § 269A (a).

may represent both a return on the individual's capital investment and a product of the individual's labor, distinguishing one from the other may be arbitrary. But any method used to make the distinction should be applied in all cases. The JCT Staff proposal uses an approach that appears to enjoy widespread appeal. However, that proposal only applies its approach in the context of flow-through business entities. The proposal described above merely extends that approach to closely held corporations, resulting in a uniform standard.

C. Qualified Personal Service Corporations

As a general rule, the income of a corporation is taxed under a system of graduated marginal rates. Under this system, the taxable income of a corporation falls into several layers, each one being subject to a (generally) higher rate than applied to the one below it. In many cases, this system will cause at least a portion of the corporation's income to be taxed at a rate as low as 15%, even if the corporation may have enough income to put it in a higher rate bracket. However, all income is taxed at the highest marginal rate that applies to individuals (currently 35%) if the corporation is a "qualified personal service corporation."¹⁹⁵ A corporation acquires that status if substantially all of its activities involve the performance of services in the fields of health, law, engineering, architecture, accounting, actuarial science, performing arts, or consulting.¹⁹⁶ The rule effectively prevents a corporation from enjoying the tax savings it would otherwise enjoy of the graduated rates were allowed to apply. It also makes it more costly for an individual to operate through a corporation than to operate as a sole proprietor, where the graduated rates would come into play.

In the absence of the rule, an individual in a high income bracket could divert a portion of his income to a corporation, where a lower tax might be imposed under the system of grad-

¹⁹⁵ I.R.C. § 11(b)(2).

¹⁹⁶ I.R.C. § 448(d)(2)(A).

uated marginal rates. A flat 35% tax on all the income of the corporation reduces or eliminates any savings that might otherwise be gained by an individual using a corporation to operate a service business. In a similar way, the modifications suggested to the JCT Staff proposal eliminate any employment tax savings that might otherwise be gained in situations where an individual operates a business through a corporation.

D. Personal Holding Company Tax

The personal holding company tax represents an instance in which the law imposes a penalty when a corporation is used in a way to minimize or avoid a tax that would otherwise apply. As explained above, the hallmark of the corporate tax system is the fact that profits are taxed first when earned by the corporation and again when distributed to shareholders. However, the shareholder level tax will not come into play if the earnings are not distributed. The personal holding company tax operates to penalize a corporation in certain situations when profits are not distributed. The amount of the penalty is equal to what the shareholders would have paid in tax.¹⁹⁷

The penalty tax can only apply if ownership of the corporation is concentrated. Specifically, more than half of the value of the corporation's stock must be owned by five or fewer individuals (and certain related parties) at any time during the last half of the tax year.¹⁹⁸ If the stock ownership test is met, the second question is whether the corporation meets an income test, which requires that at least 60% of the corporation's gross income (after certain adjustments) be passive type items, like dividends, interest, certain rents and royalties.¹⁹⁹

Because the personal holding company tax provisions consider whether stock in the corporation is concentrated, they

¹⁹⁷ I.R.C. § 541. The penalty tax is not a substitute for the tax imposed on an actual dividend received by a shareholder from the corporation. It is an additional levy. Thus, even when a corporation pays the personal holding company tax, any later dividends paid by the corporation will be taxed to the shareholder as usual.

¹⁹⁸ I.R.C. § 542(a)(2). Attribution rules operate to cause a shareholder to constructively own shares actually owned by certain related parties. I.R.C. § 544(a).

¹⁹⁹ I.R.C. §§ 542(a)(1); 543.

acknowledge the potential for closely held corporations to operate in undesirable ways that widely held corporations do not. In this case, the provisions address the risk that (primarily) high income individuals would attempt to reduce their tax bill by transferring to a corporation income producing assets. In the absence of the transfer, the income from such assets would be subject to tax at the high marginal rates that apply to the owner. However, the tax bill could be much less if the income was taxable to a corporation where lower marginal tax rates would apply. Furthermore, if the corporation never distributed the earnings, the shareholder tax that applies to dividends would never come into play.

The personal holding company tax provisions essentially make it very unattractive to use a corporation as a vehicle to separate an individual from income that ought to be taxable to him. Under the current set of rules, and under the proposal made by the JCT Staff, a corporation can operate as a vehicle that separates an individual from income that ought to be subject to employment tax as labor income. The suggestions made in this article prevent that from happening. However, they do so without operating as a penalty. Instead, the suggestions merely require all individuals in economically similar situations to be subject to the same set of rules.

E. Accumulated Earnings Tax

Even in situations where a corporation is not subject to the personal holding company tax, it may nevertheless face a penalty for failing to pay dividends to shareholders so as to prevent the shareholder level tax from applying to its earnings.²⁰⁰ Specifically, the Internal Revenue Service can assess an accumulated earnings penalty tax on any corporation that accumulates earnings beyond the reasonable needs of the business.²⁰¹ This power can be exercised only in cases where there

²⁰⁰ The tax does not apply if the following penalty taxes come into play: the personal holding company tax imposed in section 542, the foreign personal holding company tax imposed by section 552, the passive foreign investment company tax imposed by section 1297.

²⁰¹ See I.R.C. § 537 (describing how to measure the reasonable needs of the

is evidence that the corporation was utilized to avoid or evade the tax that shareholders would have to pay on dividends.²⁰²

The accumulated earnings tax represents one way to restrain the ability of a corporation to control when the shareholder level tax comes into play. The measure proposed by this article constitutes a different kind of mechanism directed at a similar activity. Under current law and under the JCT Staff proposal, the employment tax liability of a corporation's owner-employee could be controlled in undesirable ways. That power is eliminated under the measure proposed in this paper, resulting in more uniform treatment of all employee-owners whose economic situations are substantially the same.

VI. CONCLUSION

Employment taxes account for over 38% of total federal tax receipts.²⁰³ Yet the tax is assessed and collected under an outdated set of rules that is in desperate need of reform. Not only are there gaps in the law, the rules themselves offer a number of opportunities for taxpayers to arrange their affairs so as to prevent the tax from coming into play when it otherwise would.

Prior to the enactment of the temporary tax cut on dividends, the defects in the employment tax scheme were severe enough. However, after those tax cuts took effect, the defects have taken on greater significance because they appear to jeopardize the government's ability to collect the employment tax from a substantial number of people. Specifically, high income individuals who own and control corporations that employ them stand to save substantial amounts by substituting dividends for compensation they would have otherwise received from the business.

As it stands, there does not appear to be much tangible evidence that the situation will change. The JCT Staff has advanced a sensible proposal that serves as a very helpful

business).

²⁰² See I.R.C. § 533.

²⁰³ INTERNAL REVENUE SERVICE DATA BOOK 2005, PUBLICATION 55B (March 2006) tbl. 1.

framework for addressing the problem. However, there does not seem to be much interest in adopting it or any other measure directed at addressing any of the well understood defects in the employment tax laws. The last legislative measure was introduced during the Clinton administration. By contrast, there have been repeated attempts by the Bush administration and others to make permanent the dividend tax cuts, which create new and hard-to-resist opportunities for the rich to underpay employment tax by exploiting defects in the law. Those defects are already long overdue for correction. A very bad situation will only be made much, much worse if the dividend tax cuts are made permanent in the absence of legislation that eliminates the opportunity to substitute a dividend for compensation. Quite simply, the law should not permit the rich to get richer at the government's expense by working for free.

COMMENT

Tilting at Windmills: The Relationship Between Men's Non-Revenue Sports and Women's Sports

INTRODUCTION

Because they spend much of their time and energy fighting each other for scraps left over from men's revenue sports, proponents of men's non-revenue sports and women's sports are overlooking the fact that they could be valuable allies in shaping the evolution of Title IX in college athletics. The struggle between men's non-revenue sports and women's sports stems from the application of Title IX to sports. Title IX provides that "[N]o person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance."¹ While Title IX applies to all education programs and

¹ 20 U.S.C. § 1681(a) (2000). Title IX was enacted as part of the Education Amendments of 1972. SECRETARY'S COMM'N FOR OPPORTUNITY IN ATHLETICS, U.S. DEP'T OF EDUC., OPEN TO ALL: TITLE IX AT THIRTY 2 (2003), *available at* <http://www.ed.gov/about/bdscomm/list/athletics/title9report.pdf> [hereinafter OPEN TO ALL]. On June 27, 2002, Rod Paige, Secretary of the U.S. Department of Education,

has generally been well received,² its application to college athletics has been controversial.³ Even in athletics, whatever problems Title IX may or may not have, it is difficult to deny that Title IX has contributed to unprecedented participation in intercollegiate athletics by female athletes.⁴ Despite all its successes, one of the loudest and most consistent criticisms of Title IX is that it results in a drastic reduction in the number of intercollegiate athletic opportunities for males.⁵ While Title

created the Secretary's Commission on Opportunities in Athletics. *Id.* at 1. While the commission was presented with seven key questions, most of the recommendations of the commission came under the general topics of guidance and enforcement. *Id.* at 2. Reviewing the issues presented in the commission's report will give a broad overview of the issues relevant to this article.

² See OPEN TO ALL, *supra* note 1, at 2. "In 1972, when Title IX was enacted, 44 percent of all bachelor's degrees were earned by women, as compared to 57 percent in 2000. During the same period, the percentage of women earning medical degrees increased from 9 to 43 percent." *Id.* (internal citations omitted).

³ See Matthew L. Daniel, *Title IX and Gender Equity in College Athletics: How Honesty Might Avert a Crisis*, 1995 ANN. SURV. AM. L. 255, 257-59 (1995) (noting that because of controversy surrounding it, any mention of Title IX is assumed to refer to gender equity in college athletics).

⁴ See OPEN TO ALL, *supra* note 1, at 2.

Title IX (along with changes in the view of what roles girls and women should play in society) has also had a marked impact on the athletics programs offered by educational institutions across the country. The number of women's and girls' athletic teams at both the high school and college levels has dramatically increased. From 1981 to 1999, the total number of college women's teams increased by 66 percent. The growth of certain women's sports has been explosive during this period. According to the General Accounting Office, for example, colleges created over 846 new women's soccer teams. In 1971, 294,015 girls participated in high school athletics. Today, over 2.7 million girls participate in high school athletics. This represents an 847 percent increase.

Id. (internal citations omitted).

⁵ See Eric Bentley, *Title IX: The Technical Knockout for Men's Non-Revenue Sports*, 33 J.L. & EDUC. 139, 140 (2004). Between 1985 and 2001, male athletes lost 57,700 athletic opportunities at the intercollegiate level. OPEN TO ALL, *supra* note 1, at 10 (quoting Jerome Kravitz, consultant to the U.S. Department of Education and professor at Howard University). Critics respond that while some sports have been disproportionately affected, male participation in intercollegiate sports actually experienced a net gain over the same period. See Rich Haglund, *Staring Down the Elephant: College Football and Title IX Compliance*, 34 J.L. & EDUC. 439, 446 (2005) (citing NATIONAL WOMEN'S LAW CENTER, TITLE IX AND MEN'S "MINOR" SPORTS: A FALSE CONFLICT, available at <http://www.nwlc.org/pdf/>

IX is not totally to blame for the deteriorating state of men's non-revenue sports, an unfortunate side effect of the current application of Title IX is that proponents of men's and women's sports spend their energy and resources fighting each other for a shrinking piece of the athletic financial pie instead of working together to focus on eliminating true barriers to the expansion of intercollegiate-athletic opportunities for both men and women.

This paper examines the source of the controversy surrounding Title IX and athletics, including how the current application of Title IX pits proponents of men's non-revenue sports and women's sports against each other. This paper then discusses why men's non-revenue sports and women's sports are not natural enemies and how proponents of men's non-revenue sports and women's sports can work together in order to bring about reform in the application of Title IX. Finally, this paper briefly examines some of the legal obstacles to meaningful Title IX reform. By working together to lobby for a spending cap on individual revenue sports as well as an exemption of revenue-producing sports from the application of the proportionality prong and an exemption for the NCAA from antitrust laws, proponents of men's non-revenue sports and women's sports can help eliminate some of the negative effects of Title IX. The spending cap and exemptions would decrease the negative impact of Title IX on men's non-revenue sports while also ensuring the number of intercollegiate-athletic opportunities for women continues to grow.

Minor.may02.pdf). However, in some particularly troubling instances, universities have actually cut male programs that have external funding available simply to ensure compliance with Title IX, essentially eliminating opportunities for male athletes with no corresponding gain in opportunities for females. See *Chalenor v. Univ. of N.D.*, 291 F.3d 1042 (8th Cir. 2002); see also Daniel, *supra* note 3, at 286 n.166 ("Notice that Brown failed [the proportionality] prong in part because it recognized self-supported 'donor-funded' teams and permitted them to compete at the varsity level, implying incredulously that it would have been wiser for Brown not to allow all women's teams to compete when the school cannot provide funding for them." (citing *Cohen v. Brown*, 879 F. Supp. 185, 212-13 (D.R.I. 1995))).

COMPLIANCE WITH TITLE IX IN ATHLETICS

In most areas where it applies, Title IX is not very controversial. No one argues against the premise that people should not be denied educational opportunities simply because of their sex. With respect to college athletics, however, controversy arises because of the options, or lack thereof, for complying with Title IX. Courts generally recognize three ways for colleges to comply with the participation requirements of Title IX, often referred to as the "three-part test" or "three-prong test," although only one of the prongs needs to be met for a school to show compliance.⁶ The three-prong test was developed from a policy interpretation issued by the Department of Education's Office of Civil Rights (OCR) in 1979.⁷ The first prong, sometimes referred to as the "proportionality prong," requires that "a school provide[] athletic opportunities for its male and female students in numbers substantially proportionate to their respective enrollments."⁸ The second option for compliance is for a school to show "a history and continuing practice of program expansion which is demonstrably responsive to the developing interests and abilities of the underrepresented sex."⁹ The third option is for colleges to demonstrate that "the interests and abilities of the members of [the underrepresented] sex have been fully and effectively accommodated by the present program."¹⁰

⁶ See Ross A. Jurewitz, *Playing at Even Strength: Reforming Title IX Enforcement in Intercollegiate Athletics*, 8 AM. U.J. GENDER SOC. POL'Y & L. 283, 299-300 (2000).

⁷ Elisa Hatlevig, *Title IX Compliance: Looking Past the Proportionality Prong*, 12 SPORTS LAW. J. 87, 90 (2005); see Title IX 1979 Policy Interpretation, 44 Fed. Reg. 239 (Dec. 11, 1979), available at <http://www.ed.gov/about/offices/list/ocr/docs/t9interp.html> [hereinafter 1979 Policy Interpretation].

⁸ Hatlevig, *see supra* note 7, at 90 (paraphrasing 1979 Policy Interpretation).

⁹ *Id.*

¹⁰ *Id.* at 91.

THE PROPORTIONALITY PRONG AS THE ONLY PRACTICAL
OPTION FOR COMPLIANCE

In theory, the three-prong test provides flexibility to schools in complying with Title IX. While the proportionality prong is the most objective measure, the second prong allows schools time to comply as long as they are continually expanding opportunities, and the third prong allows for schools to take the different interests and abilities of their student body into account when determining how to comply with Title IX. In reality, however, practical considerations make it likely that schools will choose to satisfy the proportionality prong in order to comply.

One reason for this is that it is the most objective of the tests and satisfying the proportionality prong is more likely to prevent expensive lawsuits.¹¹ Universities' reliance on the proportionality prong to the exclusion of the other two prongs is even more understandable in light of the fact that the OCR designated the first prong a "safe harbor."¹² Also, the second and third prongs are difficult to satisfy. The second prong is by definition a temporary solution and only applies as long as the school is moving toward compliance with the proportionality prong.¹³ The third prong, while probably the most

¹¹ See Jurewitz, *supra* note 6, at 319 ("To many institutions, [cutting men's sports] . . . appears to be the easiest, most clearly defined, and most cost-effective method to comply with Title IX."). While strict compliance with the proportionality prong at the expense of certain men's non-revenue sports has caused some lawsuits, alleging reverse discrimination in violation of Title IX, none of these lawsuits have been successful at this point. See, e.g., *Chalenor v. Univ. of N.D.*, 291 F.3d 1042 (8th Cir. 2002); *Kelley v. Bd. of Trustees*, 35 F.3d 265 (7th Cir. 1994); *Harper v. Ill. State Univ.*, 35 F. Supp. 2d 1118 (C.D. Ill. 1999); *Gonyo v. Drake Univ.*, 837 F. Supp. 989 (S.D. Iowa 1993); see also *Nat'l Wrestling Coaches Ass'n v. U.S. Dep't of Educ.*, 366 F.3d 930 (D.C. Cir. 2004) (dismissing suit against U.S. Department of Education for lack of standing).

¹² See *OPEN TO ALL*, *supra* note 1, at 26. In 2003, the OCR issued a "Dear Colleague" letter removing the "safe harbor" designation from the first prong. Hatlevig, *supra* note 7, at 115.

¹³ Earl C. Dudley, Jr. & George Rutherglen, *Ironies, Inconsistencies, and Intercollegiate Athletics: Title IX, Title VII, and Statistical Evidence of Discrimination*, 1 VA. J. SPORTS & L. 177, 197 (1999) ("The second part . . . requires continuing expansion of opportunities for the 'underrepresented' sex (almost always women)

fair and equitable measure of compliance, is difficult to satisfy. In *Cohen v. Brown University*,¹⁴ the First Circuit Court of Appeals stated that "even if it can be empirically demonstrated that, at a particular time, women have less interest in sports than do men, such evidence, standing alone, cannot justify providing fewer athletic opportunities for women than for men."¹⁵ During the time relevant to *Cohen v. Brown University*,¹⁶ eight times as many male students as female students participated in Brown's intramural sports program.¹⁷ The court reasoned that disparate interest in athletics was a result of discrimination, and statistical evidence could not be used to show a lack of interest,¹⁸ virtually ensuring the only way a school could prove that it satisfied prong three was to satisfy the substantial proportionality prong.¹⁹

and it seems to assume a definition of 'underrepresentation' based on the enrollment of women in the student body as a whole."); see also *Cohen v. Brown Univ.*, 991 F.2d 888, 903 (1st Cir. 1993) (rejecting university's attempt to satisfy the second prong, despite "impressive growth" in women's athletic program, because cutting a women's program does not satisfy requirement of continuing expansion); Christopher Paul Reuscher, *Giving the Bat Back to Casey: Suggestions to Reform Title IX's Inequitable Application to Intercollegiate Athletics*, 35 AKRON L. REV. 117, 127 (2001) ("[I]t seems unlikely that an institution failing to satisfy Prong I will be able to satisfy Prong II.").

¹⁴ 101 F.3d 155 (1st Cir. 1996).

¹⁵ *Id.* at 180.

¹⁶ *Id.* at 161-64.

¹⁷ Jurewitz, *supra* note 6, at 332 (citing Appellant's Brief in Appeal from the District Court at 7, *Cohen v. Brown Univ.*, 101 F.3d 155 (1st Cir. 1996) (No. 95-2205)).

¹⁸ *Cohen*, 101 F.3d at 179-80.

¹⁹ The judicial interpretations that came out of the *Cohen* cases deserve no small part of the blame for the negative effects of Title IX on men's non-revenue sports. The court seemed to ignore the language of the Department of Education's policy interpretation and essentially struck down the third prong based on the assumption that "no person could continuously keep track and summarize 'students interests and abilities.'" Reuscher, *supra* note 13, at 133. At least one court disagrees with this assumption, stating, "[T]his Court will not join in assuming that athletic directors in this country are incapable of meeting the burden of Title IX and its regulations which incorporates a knowledge regarding their student body, effective analysis of and meeting students' needs, and filling those needs in a non-discriminatory fashion." *Pederson v. La. State Univ.*, 912 F. Supp. 892, 914 (M.D. La. 1996) (rejecting use of substantial proportionality prong as sole means of compliance).

Understandably, the reasoning of the First Circuit Court of Appeals has left many people with the impression that the third prong is no longer an option for satisfying Title IX.²⁰ Part of the problem is that there has traditionally been little guidance in how to determine if a school is complying with the third prong.²¹ While the Department of Education issued an additional clarification on March 17, 2005, allowing schools to show that they are fully meeting their female students' interest in sports simply by sending an email survey to all female students and then claiming that a failure to respond indicates a lack of interest in playing sports, questions still exist as to the effectiveness of this method.²² For one, considering the amount of spam that many university students certainly receive, attributing non-responses to surveys sent through emails as an indication of a lack of interest is a questionable tactic.²³ In addition, there is a question of what population universities should be surveying.²⁴ Should they be allowed to only survey the student body, or should they be required to survey a population representative of the area in which they recruit?²⁵ The courts have consistently recognized that inter-

²⁰ See, e.g., Juravitz, *supra* note 6, at 317-18. The OCR has issued several clarifications that are apparently intended to stress that cutting men's sports is not required to comply with Title IX and that prongs two and three are viable options. Bently, *supra* note 5, at 153 (2004). However, as long as courts continue to interpret prongs two and three out of the test, schools facing financial constraints will have to cut men's sports in order to comply with Title IX.

²¹ OPEN TO ALL, *supra* note 1, at 25-26; see Hatlevig, *supra* note 7, at 96 ("[T]here are no provisions or guidelines to help assess students' abilities and interest,' and no court has stepped forth to set such parameters." (quoting David Klinker, *Why Conforming with Title IX Hurts Men's Collegiate Sports*, 13 SETON HALL J. SPORT L. 73, 84 (2003))). Suggested ways to measure interest include "questionnaires and interviews of students, coaches, and administrators; analysis of current participation in intramural, club, or interscholastic sports; and through requests by students to add a particular sport." Mark Hammond, *Substantial Proportionality Not Required: Achieving Title IX Compliance Without Reducing Participation in Collegiate Athletics*, 87 KY. L.J. 793, 808 (1999).

²² Posting of Neena Chaudhry to ACS Blog, <http://www.acsblog.org/guest-bloggers-2699-guest-blogger-title-ix-policy-threatens-womenas-progress-in-sports.html> (March 17, 2006, 23:04 EST) [hereinafter Chaudhry].

²³ *Id.*

²⁴ Juravitz, *supra* note 6, at 288.

²⁵ B. Glenn George, *Who Plays and Who Pays: Defining Equity in Intercolle-*

est develops alongside opportunity.²⁶ If a school does not provide opportunities for female athletes, it will probably attract a female student body that is less concerned with athletic opportunity, which would lower the hurdle to comply with Title IX. Even with the Department of Education's new guidelines for gauging interest, it is nearly impossible to determine whether a lack of interest in athletic opportunities by the female student body is a result of inherent differences between males and females or a result of a lack of athletic opportunities originally available to the females of its student body.²⁷

As a result of these practical considerations, schools will generally choose to satisfy the proportionality prong. Unfortunately, because of the scarcity of resources available to college athletics, this ensures that any cuts made will apply to men's non-revenue sports.²⁸

WHY THE PROPORTIONALITY PRONG IS PROBLEMATIC

It's worth considering why so many people have problems with Title IX. Many people have trouble understanding how anything could be wrong with ensuring that an equal number of men and women get the opportunity to play college sports. The problem stems from the inherent inconsistencies in Title IX's application to college athletics.²⁹ In other areas where

giate Athletics, 1995 WIS. L. REV. 647, 656-57 (1995).

²⁶ Chaudhry, *supra* note 22; see, e.g., *Cohen v. Brown Univ.*, 101 F.3d 155, 179 (1st Cir. 1996).

²⁷ George, *supra* note 25, at 659-60; see also Chaudhry, *supra* note 22.

²⁸ David Aronberg, *Crumbling Foundations: Why Recent Judicial and Legislative Challenges to Title IX May Signal Its Demise*, 47 FLA. L. REV. 741, 811 (1995) ("[I]n reality, the proportionality requirement is the only feasible choice for universities facing budget constraints. The proportionality test is the only provision that allows schools to comply with Title IX by eliminating men's teams instead of spending unlimited amounts of money to fund women's teams.").

²⁹ See Daniel, *supra* note 3, at 291 n.190. When female plaintiffs bring Title IX claims, courts consistently ignore evidence that the whole athletic program should be considered and instead focus on the individual plaintiffs, but when men file suits, courts are suddenly unconcerned about individual plaintiffs and are instead focused on whether individual institutions are able to exercise their own judgment in how to allocate funds. *Id.*

Title IX applies, such as the classroom, opportunities are available on a gender-neutral basis. For example, there are not separate male and female engineering majors, there is simply an engineering major.³⁰ Title IX does not require that an equal number of engineering spots be open for men and women, it simply requires that no person be denied an opportunity to major in engineering because of his or her sex.³¹

With regards to athletics, however, it is exactly the opposite. Athletics is perhaps the one area where people generally agree that men and women are not on equal playing fields.³² If the NCAA were to quit designating sports as men's or women's sports, the practical effect would be to greatly reduce the number of females playing intercollegiate athletics, a result that would be unacceptable to almost everybody. While removing men's and women's designations in theory would be the least discriminatory method for enforcing Title IX (in the sense that participation would be based solely on interest and ability with no consideration of gender),³³ the practical effect would be to significantly drop the level of female participation in intercollegiate athletics, possibly to below the pre-Title IX levels. In order to ensure that males do not receive the vast

³⁰ Reuscher, *supra* note 13, at 118 n.9 ("Rev. E. William Beauchamp, the executive vice president of Notre Dame, correctly stated the intended purpose of Title IX when he said 'we don't tell our engineering department it has to have the same number of men and women, we don't do anything like that anywhere in our institution. It's based on interest.'" (quoting Ed Sherman, *New Athletic Rivalry: Men v. Women*, CHICAGO TRIBUNE, Jan. 24, 1993, at C1)).

³¹ *Id.*

³² While difficult to quantify the differences in ability between men and women at the intercollegiate level, examining the results of the NCAA Division I 2006 Indoor Track and Field Championships might give a general idea. In 2006, at least seven males in the 60 meter dash, 400 meter dash, and high jump had performances that would have broken the women's NCAA all time record in each of those events. See generally 2006 NCAA Division I Men's & Women's Indoor Track & Field Championships, NCAASPORTS.COM, <http://www.ncaasports.com/track-and-field/championship-results/indoor/2006/div1/d1> (last visited Oct. 9, 2006).

³³ One of the obstacles to reaching Title IX compromises that are acceptable to everyone is the inherent difficulty in defining gender equity. Is equity determined by equal results or equal opportunities? See Daniel, *supra* note 3, at 261 n.25. This determination is made much more difficult by the fact that athletic programs are generally segregated. *Id.*

majority of spots on intercollegiate athletic rosters and the benefits that accompany them, Title IX mandates sexual discrimination and, in practice, if not in theory, imposes a quota on participation.³⁴ While proponents of Title IX are fond of pointing out that Title IX does not require schools to cut men's non-revenue sports,³⁵ universities are left with little practical choice.³⁶ Because of the scarcity of resources for college athletics, because of the large drain "revenue sports" place on those resources, and because of the uncertainty in satisfying Title IX using anything but the proportionality prong, universities feel compelled to cut men's non-revenue sports in order to comply with the proportionality prong.³⁷

³⁴ See *supra* notes 11-28 and accompanying text; Reuscher, *supra* note 13, at 157 (recognizing that despite the original intent and legislative history of Title IX, it currently operates as a quota system and not an anti-discrimination statute).

³⁵ See, e.g., Bentley, *supra* note 5, at 145-46; Susan Eckes, *Another Pin For Women: The National Wrestling Coaches Associations' Title IX Case is Dismissed*, 182 ED. LAW REP. 683, 697 (2004). This response, if not disingenuous, is not particularly helpful to any debate on Title IX. The practical result of Title IX is that many schools feel (in large part due to courts' interpretations of the third prong) that the best way to comply is to cut men's programs, and the options theoretically available under Title IX do not lessen the negative impact on male athletics.

³⁶ See *supra* notes 11-28 and accompanying text.

³⁷ See *supra* notes 11-28 and accompanying text. For further discussion on why it is unrealistic to expect individual schools to balance the competing interests between revenue and non-revenue sports, see Gary R. Roberts, *Evaluating Gender Equity Within the Framework of Intercollegiate Athletics' Conflicting Value Systems*, 77 TUL. L. REV. 997, 999-1006 (2003). Another gutsy tactic some proponents of Title IX choose is to criticize universities for choosing to fund football over non-revenue sports. See Eckes, *supra* note 35, at 699. Even if most football revenue does not trickle down to other sports, a football team does provide a large number of opportunities for male athletes, often at no cost to the university. Choosing to cut a self-sufficient program will do nothing but decrease overall opportunity. See Daniel R. Marburger & Nancy Hogshead-Makar, *Is Title IX Really to Blame for the Decline in Intercollegiate Men's Nonrevenue Sports?*, 14 MARQ. SPORTS L. REV. 65, 81-82 (2003) (recognizing that the marginal benefit of each dollar spent on football and basketball tends to exceed the marginal benefit of each dollar spent on non-revenue sports). Another startling argument is that colleges should not be allowed to consider financial factors when determining whether to drop a sport. See Bentley, *supra* note 5, at 143. Bentley's argument apparently rests on the belief that athletic departments can always raise more money and never really addresses what options would be available when universities face a budget crunch that cannot be overcome by renewed fundraising efforts.

Another problem with the substantial proportionality prong is that it does not accurately indicate whether schools are effectively achieving the goals of Title IX.³⁸ Substantial proportionality is measured by comparing the opportunities provided to each sex with the proportion of each sex in the student body.³⁹ The student body, however, is generally not the pool from which college athletes are drawn.⁴⁰ Generally, college athletes are drawn from the pool of eligible people with the talent and interest in college athletics, which includes people from all over the globe.⁴¹ A slightly more realistic measure would be to compare the opportunities provided to each sex with the proportion of each sex involved in high school athletics in the state or region,⁴² although there would still be the problem of determining whether the disparate participation in sports was a result of past discrimination.⁴³

The effects of the proportionality prong on men's non-revenue sports, particularly those such as wrestling, swimming, and gymnastics, have been troubling.⁴⁴ From 1981 to 2001, "men lost between 1,290 to 1,434 teams and 57,100 to 57,700 participation opportunities."⁴⁵ Over the same period, women have gained around 51,967 athletic opportunities.⁴⁶

See id.

³⁸ Juravitz, *supra* note 6, at 288.

³⁹ *Id.*

⁴⁰ *Id.*

⁴¹ *Id.*

⁴² Another possible pool to be considered when applying the proportionality test is "the number of male and females that have the ability and the desire to participate in intercollegiate sports and the number of male and female athletes at the institution." Reuscher, *supra* note 13, at 147 (citing Regents of Univ. of Cal. v. Bakke, 438 U.S. 265 (1978) (rejecting in theory the substantial proportionality test for one that takes into account the most important qualification: ability to participate)).

⁴³ See *Cohen v. Brown Univ.*, 101 F.3d 155, 179 (1st Cir. 1996) ("Interest and ability rarely develop in a vacuum; they evolve as a function of opportunity and experience.").

⁴⁴ Between 1981 and 1999, wrestling lost over 170 varsity programs. See Marburger & Hogshead-Makar, *supra* note 37, at 65.

⁴⁵ See OPEN TO ALL, *supra* note 1, at 24.

⁴⁶ *Id.* While Title IX does not deserve all the blame for the lost male opportunities, when budgetary restraints make it impossible to increase the total number

When considering the fairness of such results, it is important to look at Title IX from the different value systems involved. Women's proponents may have a tendency to look at it as the total opportunities for women sports versus total opportunities for men. From this viewpoint, while the loss of athletic opportunities for men may not be desirable, it is an acceptable sacrifice in order to gain a proportional number of opportunities for women.⁴⁷ To proponents of men's non-revenue sports, however, the grouping of men's football and basketball with men's wrestling, gymnastics, and swimming is unnatural. There is very little similarity between men's revenue and non-revenue sports. Participation in men's non-revenue sports, it could be argued, is still an extra-curricular activity that enriches the participant's education, whereas for many football and basketball players, school has become an extracurricular activity that is available along with their athletic endeavors. Additionally, revenue sports are not allies of men's non-revenue sports; proponents of men's revenue sports do not take a special interest in men's non-revenue sports just because the participants are of the same sex. To proponents of revenue sports, any non-revenue sport, whether male or female, is a drain on resources that could potentially affect their competitiveness.⁴⁸ To proponents of men's non-revenue sports, they

of opportunities, compliance with Title IX does require a decrease in men's opportunities.

⁴⁷ Some authors argue that men's sports are not cut because of Title IX, but because of rising costs, lack of support, lack of success, or reallocation of resources. See Haglund, *supra* note 5, 442-43. This argument ignores the fact that it's generally only men's sports that are cut. This is troublesome to people that recognize the different values of and benefits provided by revenue-producing sports because it seems to imply that women's non-revenue sports are inherently more valuable than men's non-revenue sports.

⁴⁸ To be fair, the perspective of proponents of revenue-producing sports should be considered. It's a perfectly valid argument that the most resources should be allocated to the sports that generate the most interest and generate the most revenue. See Marburger & Hogshead-Makar, *supra* note 37, at 81-82. Big-time sports, even when not operating at a profit, produce a myriad of benefits for the school, including increasing school pride, institutional recognition, alumni support, and enrollment. See Daniel, *supra* note 3, at 298. The problem with revenue sports is that they often have been disconnected from the educational mission of

are disproportionately suffering the consequences of the arm's race in big-time sports solely because of their gender, a result that seems to fly in the face of the purpose of Title IX.⁴⁹

As it stands now, Title IX will forever pit proponents of women's sports against proponents of men's non-revenue sports.⁵⁰ Applying the proportionality prong will protect women's progress but will always result in a disproportionately negative effect on men's non-revenue sports; doing away with the proportionality prong or increasing the use of the accommodating interests prong will ease the burden on men's non-revenue sports but will threaten to slow the progress of women in intercollegiate sports.

AN ALTERNATIVE APPROACH, WOMEN'S AND MEN'S NON-REVENUE SPORTS AS ALLIES

While the current political climate pits women's sports versus men's non-revenue sports, this result is not inevitable. A different result could be had if proponents of women's sports

the institution. *See id.* at 299. Other than the few athletes who get training to pursue a career in professional athletics, the athletes get few of the benefits they produce.

⁴⁹ *See Bentley, supra* note 5, at 140 ("It seems perplexing that a statute, which clearly provides that no person shall be denied participation in an educational program 'on the basis of sex,' would allow an institution to drop a men's program to become compliant with Title IX, because by doing so, the institution seems to deny a person (a male athlete) participation in an educational program (intercollegiate athletics) on the basis of their sex (being male)."). Discontinuing men's non-revenue sports and not women's non-revenue sports seems even more inequitable if one believes that football is not designed to provide benefits to only male students, but rather to provide benefits to male and female students alike by acting as a "profit center" for the university. Daniel, *supra* note 3, at 292. Not only are the individual males participating in the non-revenue sports negatively effected by the operation of revenue sports, males in general are not provided an equal number of opportunities in sports that are still connected to the educational mission of the university. *See Daniel, supra* note 3, at 308 (recognizing that current enforcement of Title IX leaves men with "a dearth of 'educational' sports, a situation which [is] as unfair to men as the past situation was to women.").

⁵⁰ Haglund, *supra* note 5, at 447. Title IX has created a battle for scarce resources between male and female athletes, which "has led to distrust and closed communications between partisan groups seeking to further the interests of their members." Juravitz, *supra* note 6, at 287.

and men's non-revenue sports considered themselves political allies. While this would require a paradigm shift, it's not as drastic as it first sounds.⁵¹

First, it should be recognized that men's non-revenue sports and women's sports share more in common as non-revenue-producing sports⁵² than men's non-revenue sports and men's revenue sports share in common as sports with participants of the same gender. Revenue sports are no longer extracurricular activities that enhance student-athletes' educational experiences; they are "entertainment product[s] sold to the public through tickets and broadcasting rights for billions of dollars."⁵³ The largest benefit of revenue sports is not to provide educational and enrichment opportunities to its participants,⁵⁴ but to increase "student and alumni cohesion."⁵⁵ The NCAA itself, at least with respect to Division I, recognizes that some of its sports are not necessarily focused on providing benefits to the students.⁵⁶

⁵¹ The first step would be for each group to admit that each side of the debate is reasonable, depending on the value system that underlies it. Roberts, *supra* note 37, at 110 ("[F]ew who have taken ardent positions in the Title IX/gender-equity debate have been willing to acknowledge that their position and agenda derive from, and depend upon, their particular perspective (if not self-interest) and the value system that underlies it.").

⁵² While women's basketball shows promise as a revenue-producing sport and may be starting to show some of the negative effects when the business interests of college sports collide with the educational interests, for simplicity this discussion will focus on the majority of women's sports, which do not generate large revenues. See Marburger & Hogshead-Makar, *supra* note 37, at 79 (recognizing women's basketball as the only women's sport with a realistic hope of generating a profit).

⁵³ See Roberts, *supra* note 37, at 1001.

⁵⁴ Daniel, *supra* note 3, at 259; see also Reuscher, *supra* note 13, at 151 ("In reality, the purpose of athletics is not to provide lifelong skills for the student, but to operate as a 'profit center' for the institution.").

⁵⁵ Daniel, *supra* note 3, at 303; see also Marburger & Hogshead-Makar, *supra* note 37, at 75 ("According to the utilitarian perspective, college athletics garner local and/or national support for the university, generate revenue to support the university mission, and create a sense of community among students, faculty, and alumni."). As a result, the athletic department is often "autonomous and divorced from the academic structure . . ." Daniel, *supra* note 3, at 298-99.

⁵⁶ See Marburger & Hogshead-Makar, *supra* note 37, at 76.

[T]he Division I philosophy 'recognizes the dual objective in its athletics program of serving both the university or college community (participants, student body, faculty-staff, alumni) and the general public (community, area, state, nation).' More specifically emphasizing football and basketball, Division I 'sponsors at the highest feasible level of intercollegiate competition one or both of the traditional spectator-oriented, income-producing sports of football and basketball.' Further, Division I membership 'strives to finance its athletics program insofar as possible from revenues generated by the program itself.'⁵⁷

Division I policy specifically mentions the income produced by football and basketball and promotes the idea that income from those sports and others should be used, to the extent possible, to self-finance individual athletic departments.⁵⁸

Non-revenue sports of both genders, however, are more likely to be modeled after the amateurism model. Sports falling under the amateurism model are designed for the benefit of the student.⁵⁹ Amateur sports are intended to contribute to the development of the discipline, teamwork, self-confidence, and leadership skills of its participants.⁶⁰ Because non-revenue sports come closer to satisfying the amateurism model,⁶¹ they

⁵⁷ *Id.* (internal citations omitted) (quoting NAT'L COLLEGIATE ATHELETIC ASS'N, 2003-04 NCAA DIVISION I MANUAL art. 20.9).

⁵⁸ Marburger & Hogshead-Makar, *supra* note 37, at 77.

⁵⁹ Daniel, *supra* note 3, at 295. Daniel describes two extreme models for college athletics, sports that fall under the ideals of "pure amateurism" and sports that operate as "profit centers." *Id.* However, Daniel notes that most sports fall somewhere in between these two extremes. *Id.*; see also Marburger & Hogshead-Makar, *supra* note 37, at 75 (making the same distinction but referring to the pure amateurism model as the consequentialist perspective and the profit-center model as the utilitarian perspective).

⁶⁰ Daniel, *supra* note 3, at 295; see also Marburger & Hogshead-Makar, *supra* note 37, at 75 ("[A]thletics is [sic] consistent with the mission of a university because it develops desirable traits 'such as courage, integrity and coolness under pressure.'" (quoting Robert Simon, *Intercollegiate Athletics: Do They Belong on Campus?*, in *SPORTS AND SOCIAL VALUES* 126 (1985), reprinted in *RETHINKING COLLEGE ATHLETICS* 43 (Judith Andre & David N. James eds., 1991)).

⁶¹ Although the tendency has been for college athletics to consume more and more of participants' time in order for them to stay competitive, there is less

tend to be more connected to the school's educational mission.⁶²

Second, more so than competing with each other, women's sports and men's non-revenue sports are both competing with revenue sports for resources. The intense pressure to field winning teams in revenue-producing sports causes escalating spending that prevents schools from fielding new women's sports or maintaining the current men's non-revenue sports.⁶³ Although women's sports are in a much stronger position politically because of the current application of Title IX, they should not discount the value of men's non-revenue sports as political allies. For one, the superior political position of women's sports is not guaranteed. With the changing composition of the federal courts,⁶⁴ there is no guarantee that the Department of Education's policy will continue to be given deference and interpreted the same way. Also, while it's unclear how many schools would be willing to take advantage of it, the Department of Education's clarification issued on March 17, 2005 could make it feasible for schools to move away from the proportionality

pressure on the institution to make accommodations for athletes in non-revenue sports at the expense of the institution's educational goals.

⁶² Ironically, the NCAA, arguably the organization that exploits student-athletes the most, publicly promotes the amateurism model and asserts that student-athletes should be protected from exploitation by professional and commercial enterprises. See generally WALTER BYERS, UNSPORTSMANLIKE CONDUCT: EXPLOITING COLLEGE ATHLETES (1995). See also Marburger & Hogshead-Makar, *supra* note 37, at 66 ("Large athletic departments are currently using the National Collegiate Athletic Association (NCAA) as an agent to procure monopoly revenues while benefiting from the services of unpaid athletes.").

⁶³ Hammond, *supra* note 21, at 811 (1999).

⁶⁴ See generally LEADERSHIP CONFERENCE ON CIVIL RIGHTS EDUC. FUND, TURNING RIGHT: JUDICIAL SELECTION AND THE POLITICS OF POWER (2004), available at http://www.civilrights.org/publications/reports/judges/judges_report.pdf (discussing the political consequences of a judiciary branch increasingly filled with Republican nominations). While this report by the Leadership Conference on Civil Rights Education Fund provides a particularly cynical view of Republican presidents and their judicial nominees, it does support the common sense notion that as five of the last seven presidential terms have been filled by Republicans, the political orientation of judicial appointments has tended to trend to the right. See *id.*

prong.⁶⁵ If schools do take advantage of the Department of Education's clarification, there are methods to measure interest that could drastically reduce schools' obligations under Title IX. For example, if a university used participation in intramural sports as an indicator of interest, a school could offer a significantly larger number of opportunities to male athletes and still be considered to be accommodating the interests of its student body.⁶⁶ Such a development could put women's sports in the same place that men's non-revenue sports occupy now, scrambling to maintain the current number of athletic opportunities.

Even if women's sports maintain their superior political clout, there is a potential ceiling on the number of athletic opportunities available to women. If the pressures of competing in revenue sports continues to increase, more schools could move in the direction of Tulane, which only maintains six male athletic programs,⁶⁷ the minimum number of programs required to remain in Division I.⁶⁸ If other schools adopt this model to maximize the amount of resources available to revenue-producing sports, women's opportunities will be capped at the number required to match the number of male athletes in just six sports. This result would allow women to achieve substantial proportionality, but it would be a Pyrrhic victory as there would be virtually no chance for expanding the intercollegiate athletic opportunities available to women.

Also, women's sports could eventually find themselves with the same conflicts between revenue and non-revenue-producing

⁶⁵ See *supra* text accompanying notes 22-26.

⁶⁶ See Jurewitz, *supra* note 6, at 332 (citing Appellant's Brief in Appeal from the District Court at 7, Cohen II, 101 F.3d 155 (1st Cir. 1996)). Male students tend to take advantage of intramural opportunities at a much higher rate than female students, possibly indicating that male students really do have more interest in athletics than female students. See Jurewitz, *supra* note 6, at 332 (citing Appellant's Brief in Appeal from the District Court at 7, Cohen II, 101 F.3d 155 (1st Cir. 1996)).

⁶⁷ Tulane maintains football, basketball, and baseball (revenue sports), as well as cross-country, golf, and tennis (sports that cost very little to operate). Roberts, *supra* note 37, at 1006.

⁶⁸ *Id.*; see also Marburger & Hogshead-Makar, *supra* note 37, at 66 (noting that Division I universities have financial incentives to reduce the number of non-revenue sports to minimum requirement).

sports. If more women's basketball programs become revenue producers, they too may be caught up in the "arms race."⁶⁹ If this happens, schools will be able to spend disproportionate amounts of money on the "revenue-producing" women's sports, still satisfy Title IX, and leave the women's non-revenue-producing sports on a bare-bones budget. Women's sports will then end up in the same predicament as men's sports are in now, with the vast majority of resources and benefits of intercollegiate athletics going to the revenue-producing "haves" while the non-revenue-producing "have-nots" are left behind.

If, however, proponents of women's sports and men's non-revenue sports aligned their interests now, they might be able to procure advantages that will help continually increase the athletic opportunities available to both men and women while protecting sports that are still connected to the educational mission of the university.⁷⁰

WHAT THEY SHOULD LOBBY FOR

Proponents of women's sports and men's non-revenue sports should collaborate with members of the NCAA and Congress to enact comprehensive reform that includes an amendment to Title IX exempting revenue-producing sports from the formulation of the proportionality prong and an amendment to antitrust laws that would give the NCAA an exception allowing it to impose some sort of spending cap on individual college sports.

⁶⁹ This is potentially already happening, as women's basketball in Division I-A experienced an average increase of \$660,000.00 per school from 1993 and 2001, just \$220,000.00 less than their male counterparts and more than three times as much as the average men's non-revenue sport at the Division I-A level. Marburger & Hogshead-Makar, *supra* note 37, at 86-87.

⁷⁰ See Marburger & Hogshead-Makar, *supra* note 37, at 69 (noting that Title IX in its current form does not prevent schools from "abandoning the educational mission of athletics").

THE CASE FOR A SPENDING CAP

The foremost goal for proponents of women's sports and men's non-revenue sports should be a cap on the amount of money spent on revenue sports. Without a cap, any other reforms lessening the negative impact of Title IX on men's non-revenue sports would simply allow revenue sports to gain a larger piece of the pie.⁷¹ A cap would immediately benefit non-revenue sports by slowing down the "arms race"⁷² among the big time sports. Even with less money being invested, college football and basketball would likely still generate huge revenues.⁷³ If universities were limited in how much of that money they could pour back into the revenue sports, they would have no choice except to spread it to other areas of the university, one such area being non-revenue sports.

It's worth taking a moment to examine the possible options

⁷¹ See Jurewitz, *supra* note 6, at 337-38 (noting potential that exemption for revenue-producing sports would simply result in schools over-funding football and men's basketball while under-funding women's sports and men's non-revenue sports); Marburger & Hogshead-Makar, *supra* note 37, at 65-66 ("Weakening gender equity laws will only make shifting resources from the minor sports to men's football and basketball easier.").

⁷² The "arms race" refers to the escalating amounts of money athletic departments spend in an effort to stay competitive. See Kevin J. Rapp, *Forced to Punt: How the Bowl Championship Series and the Intercollegiate Arms Race Negatively Impact the Policy Objectives of Title IX*, 80 IND. L.J. 1167, 1179-80 (2005).

⁷³ The reason is that many of the expenditures in athletics do not so much increase revenues for college athletics as a whole, but rather determine how existing revenues are distributed to individual schools. See Marburger & Hogshead-Makar, *supra* note 37, at 85-86 (explaining that the profitability of a program is in part related to the successfulness of the program, and as a result college athletics are in some ways a zero-sum game, where each team increases expenditures to increase competitiveness, but does not actually increase the number of wins available). Also, college athletics seem to be at least somewhat shielded from market forces. For example, the NCAA men's basketball tournament continues to bring in more and more money, despite early entries into the NBA draft depleting the talent pool. See, e.g., *CBS Renews NCAA B'Ball*, CNN MONEY, Nov. 18, 1999, <http://money.cnn.com/1999/11/18/news/ncaa> (reporting the NCAA's new \$6 billion contract with CBS for the rights to broadcast the men's NCAA basketball tournament, an amount that provides more than twice the annual payment of the previous contract). Similarly, any decreases in the competitive level of intercollegiate athletics brought about by less than state-of-the-art locker rooms, weight-rooms, etc., would probably not significantly impact revenues.

with regard to a spending cap.⁷⁴ The most obvious option is a hard cap⁷⁵ on the amount of money that could be spent on an individual sport. This would be easiest to employ and enforce, but would be politically difficult to institute and would possibly have some undesirable effects on revenue. A hard cap might prevent institutions from investing in facilities that would produce revenue, such as stadium expansions.⁷⁶ There is also the possibility that when faced with choosing between expenditures that will benefit student athletes and expenditures that will make boosters happy, athletic departments will choose to satisfy boosters.⁷⁷ Accordingly, a soft cap would probably be more beneficial. While it might open avenues for abuse, a soft cap with exemptions for desirable expenditures, such as those on educational benefits for the players, would probably be the superior policy.⁷⁸

⁷⁴ Spending caps have been proposed by various people, but no feasible method for imposing them has been offered. See, e.g., Hammond, *supra* note 21, at 812.

⁷⁵ "Hard" refers to a cap without exceptions whereas "soft" refers to a cap with desirable exemptions for spending.

⁷⁶ Spending large amounts of money on stadium expansions is controversial to some people, but stadium expansions are mentioned not to suggest that they are always a good investment, but rather to suggest that some stadium expansions are profitable. Almost any expansion where there is a demand is more economically beneficial than, for example, spending money on leather couches to go in locker rooms or building multimillion dollar practice facilities that service one sport.

⁷⁷ For example, to ensure that universities do not choose to install nicer carpet in the stadium luxury suites instead of hiring more tutors for athletes, expenditures that are clearly educational in nature could be exempt. Likewise, if a hard, rather than soft cap were enacted, it is theoretically possible that universities could choose to comply by cutting down on the number of scholarships for football players. This is an action that many advocates of women's sports and men's non-revenue sports support. See, e.g., Eckes, *supra* note 35, at 701; Haglund, *supra* note 5, at 442-43 (2005); Reuscher, *supra* note 13, at 154-55. However, besides the obvious undesirable effect of cutting athletic opportunities in order to comply with Title IX, which is the same ill that the current version of Title IX causes, such an action could be particularly harmful to minority students. See Roberts, *supra* note 37, at 1005.

⁷⁸ Another alternative to a spending cap would be some sort of revenue sharing requirement for major sports. Such a policy would require that a certain percentage of each dollar earned by a revenue sport be spent on something other than a revenue sport. However, such a system would be difficult to structure

AMASSING SUPPORT FOR A SPENDING CAP

Unfortunately, getting a spending cap will be the most difficult goal to achieve. The most direct method of instituting a spending cap would be to have the NCAA adopt one in its by-laws.⁷⁹ A spending cap applicable to all Division I schools would be considered a division dominant proceeding and would require a 2/3 majority vote of all delegates present and voting at the division's annual or special convention.⁸⁰ In order to gather enough support to institute a spending cap, compromises must be made in order to appeal to the broadest audience possible.

In order to cement the alliance between women's sports and men's non-revenue sports, proponents of women's sports should be willing to entertain the idea of discounting athletes in revenue sports when determining substantial proportionality.⁸¹ While a spending cap on revenue sports would probably

because of the "non-profit" nature of athletic departments and the vast differences in the revenue generating potential of different schools. If defined in terms of profits, athletic programs could avoid sharing any revenues by committing virtually all revenue to capital expenses for the football program. However, if simply defined in terms of revenue, a very complex structure would be needed to account for the economic differences between the "haves," who may produce millions of dollars in profits, and the "have-nots," who may be struggling to break even.

⁷⁹ However, enacting a spending cap without a statutory exemption from the Sherman Act would raise potential antitrust problems. See *infra* notes 88-96 and accompanying text.

⁸⁰ NAT'L COLLEGIATE ATHLETIC ASS'N, 2005-06 NCAA DIVISION I MANUAL § 5.02.1.1.1, available at http://www.ncaa.org/library/membership/division_i_manual/2005-06/2005-06_d1_manual.pdf.

⁸¹ This exemption would apply to athletes in any revenue sport, regardless of gender, so female athletes at certain schools might benefit from the exemption of members of the women's basketball team. The idea of exempting revenue sports is not a new one. During the debates over the Education Amendments of 1974, the Senate actually passed the Tower Amendment, which exempted revenue sports. Jurewitz, *supra* note 6, at 292-95. The Tower Amendment did not make it out of conference, however. *Id.* Success for a similar amendment will probably be dependant on support from some women's groups that traditionally have resisted such efforts. For a brief overview of the failed Tower Amendment, and subsequently enacted Javitz Amendment, see *Id.* at 292-95. Similar attempts to exempt revenue-producing sports include the Helms Amendment in 1975 and the McClure Amendment in 1976. See Daniel, *supra* note 3, at 310 (suggesting that "these attempts failed because it was feared that 'revenue-producing' was merely a eu-

be in the best interest of men's non-revenue sports regardless of the formulation of the proportionality prong, the possible exemption of athletes from revenue sports would provide motivation for proponents of men's non-revenue sports to focus on working with proponents of women's sports to gain ground and resources from revenue sports, rather than on continuing to fight with women's sports for dwindling resources.⁸²

While men's non-revenue sports may not bring the political capital to the table that women's sports possess, proponents of women's sports could gain significant benefits by including them in their efforts. First, an alliance between proponents of women's sports and men's non-revenue sports would help illustrate that the interests of non-revenue sports are different from those of revenue sports, which generally conflict with overall goals of the educational institution.⁸³ Distinguishing the values of non-revenue sports from the values of revenue sports will be necessary to amass the support needed to take on the strong financial interests supporting the current system.⁸⁴ If non-revenue sports are not fundamentally different from revenue sports, it is too difficult to counter the argument that individual sports should be entitled to spend the money they generate. Also, many people may feel a spending cap on reve-

phism for 'desirable high-profile men's sports').

⁸² Another advantage of this solution would be to ensure that the number of scholarships for sports such as football is not reduced. Such reductions do have support from some sources who believe that the number of football scholarships and roster spots should be reduced in order to free up scholarship money and opportunities for men's non-revenue sports. See, e.g., Eckes, *supra* note 35, at 701; Haglund, *supra* note 5, at 442-43; Reuscher, *supra* note 13, at 154-55. However, while this may help ease the burden placed on men's non-revenue sports, it may be an unwise policy considering the disparate impact it would have on African-American athletes. See Roberts, *supra* note 37, at 1005. The racial impact of football scholarships is even more important in light of the argument of some scholars that universities often tend to add women's sports that are not played by inner city African-Americans. See Reuscher, *supra* note 13, at 144.

⁸³ See *supra* notes 52-57 and accompanying text.

⁸⁴ The monopolistic advantages that the NCAA's amateurism rules create protect the revenues of the NCAA. Daniel, *supra* note 3, at 301. Because of the large amount of money at stake, any reform that might threaten the NCAA's revenue sources, such as capping spending on revenue sports, will likely be met with stiff resistance.

nue sports runs counter to American capitalistic traditions of valuing performance. Recognizing that athletes from revenue sports have a different value than athletes from non-revenue sports might appeal to this section of the population.

The most important step for women's sports and men's non-revenue sports in lobbying for a spending cap will be getting individual institutions within the NCAA to buy into it. For this to happen, institutions will have to be convinced that a spending cap benefits them in some way. In reality, the majority of institutions will benefit from a spending cap. While college sports generate a lot of revenue, that revenue is still distributed mainly to dominant athletic programs.⁸⁵ Because the NCAA distributes money in part based on the success of a school's athletic teams, the gap between the "haves" and "have-nots" is difficult to close.⁸⁶ A spending cap, much like the salary cap in the NFL, will help provide parity within college football. As schools become more competitive across the board, schools that were previously unsuccessful will reap some of the benefits that accompany a winning program in major college sports, such as increased publicity and alumni giving. While some of the powerhouse programs might decline as other schools become more competitive, it is unlikely that donors/boosters will abandon their programs simply because they are not as dominant as they once were.⁸⁷

⁸⁵ From 1996 to 1997, despite all the revenue generated by men's college athletics, 62.75% of men's athletic programs still lost money. Jurewitz, *supra* note 6, at 337 n.353. However, the big-name colleges fared much better. The same year, the men's athletic department at the University of Alabama at Tuscaloosa made a profit of over \$13 million, with the men's and women's athletic program making a profit of \$8 million overall. Jurewitz, *supra* note 6, at 337 n.354; *see also*, Rapp, *supra* note 71, at 1168 (noting that schools from non-BCS conferences lose an average of \$1 million each year on their football programs).

⁸⁶ *See* Daniel, *supra* note 3, at 300-01; Marburger & Hogshead-Makar, *supra* note 37, at 89 ("The result is an 'arms race' through which the rich get richer and the poor get poorer.").

⁸⁷ The playing field in NCAA football has already been leveled several times. In 1977, the NCAA placed a cap on the number of football scholarships a school could offer at ninety-five. *Federal Judge Says NCAA Challenge Should Go to Trial*, CBS SPORTSLINE, Sept. 16, 2005, <http://cbs.sportslines.com/collegefootball/story/8853606/rss>. The number of available football scholarships was again reduced to

Legal Obstacles to a Spending Cap

As a type of horizontal restraint, a cap on individual sports does raise the possibility of antitrust challenges. The ideal solution from the NCAA's point of view would be a statutory exemption from antitrust laws. A statutory exemption would protect the NCAA from costly violations of antitrust laws and make it much more likely that the NCAA would voluntarily impose a spending cap.⁸⁸ In order to encourage the enactment of a spending cap, proponents of women's sports and men's non-revenue sports should lobby for an antitrust exemption along with a reformulation of the proportionality prong under Title IX.

Even if a statutory exemption is not granted, however, it is still possible that the NCAA could impose a spending cap on individual sports without violating antitrust laws. Courts have interpreted the Sherman Act of 1890 to prohibit unreasonable restraints of trade.⁸⁹ While certain categories of trade restraint, such as horizontal price fixing, vertical price maintenance, group boycotts, tying arrangements, and horizontal market divisions are generally considered to be per se unreasonable,⁹⁰ the Supreme Court has applied the "rule of reason" when considering horizontal restraints and the NCAA because some horizontal restraints are necessary in order for intercollegiate athletics to be available as a product.⁹¹

Under the rule of reason, the test is whether the challenged agreement is one that promotes competition or sup-

92 in 1992, 88 in 1993, and 85 in 1994. *Id.* While the aggregate number of talented players on elite teams may have decreased, overall parity between teams has increased, and NCAA football has not suffered any decline in popularity.

⁸⁸ The NCAA's attempt at limiting the salaries of restricted earnings coaches violated the Sherman Act and eventually cost the NCAA over \$54 million. Marc R. Leduc, *Law v. National Collegiate Athletic Association: A Guide to How Courts Will Treat Future Antitrust Challenges to NCAA Regulations*, 26 J.C. & U.L. 139, 167-68 (1999).

⁸⁹ *Id.* at 146; *NCAA v. Bd. of Regents of the Univ. of Okla.*, 468 U.S. 85, 98 (1984).

⁹⁰ Leduc, *supra* note 88, at 146-47.

⁹¹ *NCAA v. Bd. of Regents of the Univ. of Okla.*, 48 U.S. 85, 100-01 (1984).

presses competition.⁹² A spending cap on individual sports would promote competition by ensuring the economic viability of its members' athletic departments. However, by limiting the resources that could be spent, a spending cap might restrain competition in other respects, such as the salaries paid to coaches and support staff and the benefits provided to players. The limitation on the money spent on players would be less problematic than the limitation on salaries of coaches.⁹³ The NCAA already restricts the number and value of scholarships available, and a spending cap would not likely affect the analysis. The effects a spending cap would have on the compensation of coaches should probably cause more concern for the NCAA. In *Law v. NCAA*,⁹⁴ the NCAA was found to have violated the Sherman Act by placing a limit on the compensation of restricted earnings coaches⁹⁵ and eventually had to pay 54.5 million dollars in settlements.⁹⁶ *Law v. NCAA* would be distinguishable from a case involving a general spending cap because there would not be a direct link between the spending cap and any "material" needed to produce the NCAA's "product." Regardless of whether a spending cap would eventually survive an antitrust challenge, obtaining a statutory exemption and avoiding an antitrust challenge would likely be needed in order to convince the NCAA to adopt a spending cap.

CONCLUSION

Title IX has undoubtedly helped increase the number of intercollegiate athletic opportunities available to women. While the current enforcement of Title IX will ensure that female

⁹² *Bd. of Trade v. United States*, 246 U.S. 231, 244 (1918).

⁹³ The NCAA has already enjoyed some success in defending antitrust actions against lawsuits brought by students. See, e.g., *McCormick v. NCAA*, 845 F.2d 1338 (5th Cir. 1988) (holding that "no-draft" rules have primarily pro-competitive effects); *Gaines v. NCAA*, 746 F. Supp. 738, 744-46 (M.D. Tenn. 1990) (holding that antitrust laws do not apply to NCAA eligibility rules).

⁹⁴ 134 F.3d 1010 (10th Cir. 1998).

⁹⁵ *Id.* at 1024.

⁹⁶ Stephanie M. Greene, *Regulating the NCAA: Making the Calls Under the Sherman Antitrust Act and Title IX*, 52 ME. L. REV. 81, 83 (2000).

athletes will be partly sheltered from the "arms race" in men's college basketball and football, it is incapable of producing continual growth in intercollegiate opportunities for women and ensures that opportunities for intercollegiate opportunities in men's non-revenue sports will continue to decrease. Title IX ensures that women's programs receive a disproportionate amount of the resources left over from revenue-producing sports and ensures that they have a proportional number of scholarships but does not help them gain a status that is truly equal to that of men's sports. Proponents of women's athletics and men's non-revenue sports should align their interests and, along with the less competitive colleges in NCAA Division I, push for reform that will help protect non-revenue sports for each gender, level the playing field between the haves and have-nots in Division I-A, and help expand the overall number of intercollegiate athletic opportunities available to student athletes of both genders.

While this sort of compromise will be difficult because it is not in the interests of big time college football programs, the time for it is now. Because interest develops alongside of opportunity,⁹⁷ discontinuing men's non-revenue sports en masse ensures that there probably will not be as much interest in men's wrestling, gymnastics, swimming, and other non-revenue sports in the future, and women's sports will have lost a potentially valuable ally in their fight for equality. When this interest no longer exists, men's programs will probably consist of football, basketball, baseball, and a handful of non-revenue sports, and Title IX will never fully achieve its true purpose in college athletics, which is to end sexual discrimination and help expand women's opportunities.

Ronnie Wade Robertson

⁹⁷ See *Cohen v. Brown Univ.*, 101 F.3d 155, 179 (1st Cir. 1996) ("Interest and ability rarely develop in a vacuum; they evolve as a function of opportunity and experience.").

RECENT DECISION

Securities Fraud—Federal Preemption—SLUSA Preempts State-Law Class Actions Brought by Holders of Covered Securities*

I. FACTS

Congress passed the Securities Litigation Uniform Standards Act of 1998 (SLUSA) in an effort to curtail the unintended shift of class-action securities fraud cases from federal court to state courts in the wake of the 1995 Reform Act.¹ SLUSA's core provision prohibited private parties from bringing covered class actions based on state law in either state or federal court when such actions alleged fraud "in connection with the purchase or sale" of covered securities.² Respondent

* Outstanding Case Note Award presented by Heilman Kennedy Graham, P.A., Jackson, Mississippi.

¹ H.R. REP. NO. 105-640, at 10 (1998); S. REP. NO. 105-182, at 3-4 (1998).

² 15 U.S.C. § 78bb(f)(1) (2000). SLUSA states, in pertinent part:

(1) Class action limitations. No covered class action based upon the statutory or common law of any State or subdivision thereof may be maintained in any State or Federal court by any private party alleging—

(A) a misrepresentation or omission of a material fact in connection

Shadi Dabit, a former Merrill Lynch stockbroker, filed a state-law class action in the U.S. District Court for the Western District of Oklahoma on behalf of himself and other brokers.³ The complaint alleged that petitioner Merrill Lynch breached various duties owed to its brokers by disseminating misleading research in an effort to manipulate stock prices.⁴ In its original form, the complaint sought damages for claims relating both to the wrongfully-induced purchasing of stocks and the wrongfully-induced holding of stocks.⁵ The district court

with the purchase or sale of a covered security; or

(B) that the defendant used or employed any manipulative or deceptive device or contrivance in connection with the purchase or sale of a covered security.

15 U.S.C. § 78bb(f)(1) (2000). A "covered class action" is defined as any single lawsuit seeking damages on behalf of more than fifty people. *Id.* § 78bb(f)(5)(B). A "covered security" is defined as any security traded on a national exchange. *Id.* §§ 77r(b), 78bb(f)(5)(E). A separate provision of the statute renders all "covered class actions" filed in state court removable to federal court. *Id.* § 78bb(f)(2).

³ *Merrill Lynch, Pierce, Fenner & Smith, Inc. v. Dabit*, 126 S. Ct. 1503, 1507 (2006). Dabit filed his complaint in district court asserting the court's diversity jurisdiction, yet his claims were grounded in Oklahoma state law. *Id.* The class of plaintiffs consisted of current and former Merrill Lynch brokers who had purchased for themselves and their clients certain covered securities between December 1, 1999, and December 31, 2000. *Id.* The lawsuit was one of many that followed a 2002 investigation by New York Attorney General Eliot Spitzer into Merrill Lynch's research practices and the integrity of its analysts' stock recommendations. See *id.*; Ruth Simon, *Not All Merrill Investors Have Cases—Settlement Likely to Benefit Those Who Can Show They Bought on Analysts' Advice*, WALL ST. J., May 22, 2002, at D2. The firm subsequently settled the dispute with New York and other states for the sum of \$100 million. Simon, *supra*.

⁴ *Dabit*, 126 S. Ct. at 1507. Specifically, the complaint alleged that Merrill Lynch breached the fiduciary duty and covenant of good faith and fair dealing it owed to its brokers. *Id.* Dabit theorized that management directed the firm's research analysts to issue overly bullish appraisals of stock values in an effort to drive up the prices of stocks held by its investment banking clients. *Id.* Dabit alleged that he and other brokers relied on the biased research in deciding whether to sell or hold the affected securities and in advising their clients to that end. *Id.* In reliance on the allegedly overly optimistic research, both the brokers and their clients continued to hold stocks well past the time they would have sold had they not allegedly been misled. *Id.*

⁵ *Id.* at 1508. Dabit asserted that Merrill Lynch's alleged misconduct damaged the class members by (1) causing losses to their personal portfolio values when the prices of the affected stocks plummeted after the truth was revealed and (2) inflicting business losses on the brokers (e.g., lost commissions, fees, etc.)

in Oklahoma dismissed the complaint with leave to amend, finding that while the "purchasing" claims were certainly preempted by SLUSA, the "holding" claims might not be.⁶ Dabit re-filed an amended complaint, which was subsequently transferred to the U.S. District Court for the Southern District of New York for consolidated pretrial proceedings.⁷ The district court in New York granted Merrill Lynch's motion to dismiss, ruling that SLUSA preempted all of Dabit's claims.⁸

On appeal, the U.S. Court of Appeals for the Second Circuit vacated the judgment and remanded, holding that respondents' claims as "holders" escaped SLUSA's preemptive sweep.⁹ The court concluded that fraud is "in connection with the purchase or sale" of securities, in the context of SLUSA preemption, only if it is alleged by a purchaser or seller of securities.¹⁰ The court reasoned that as the respondents were fraudulently induced not to purchase or sell securities, but merely to "hold" (i.e., delay selling) securities, SLUSA did not preempt their claims.¹¹ Three months later, the U.S. Court of Appeals for the Seventh Circuit, considering the same issue, oppositely held that SLUSA's preemptive scope does in fact

when clients, now cognizant of their own losses, took their business elsewhere. *Id.* at 1507.

⁶ *Id.* at 1508.

⁷ *Id.* During this time period, dozens of similar suits had been filed against Merrill Lynch in the wake of the New York investigation. *Id.* See also *supra* note 3 and accompanying text. In the interest of judicial economy, the Judicial Panel on Multidistrict Litigation transferred all such suits, including Dabit's, to the District Court for the Southern District of New York. *Dabit*, 126 S. Ct. at 1508.

⁸ *Dabit*, 126 S. Ct. at 1508. The district court's opinion dismissing the action may be found at *In re Merrill Lynch & Co., Inc.*, No. 02MDL1484(MP), 2003 WL 1872820, at *1 (S.D.N.Y. Apr. 10, 2003).

⁹ *Dabit*, 126 S. Ct. at 1508. In reaching its decision, however, the Second Circuit did acknowledge the Court's prior broad interpretation of the "in connection with" phrase. *Id.*

¹⁰ *Id.* (citing 15 U.S.C. § 78bb(f)(1) (2000)).

¹¹ *Id.* The Second Circuit remanded the case to be dismissed without prejudice. *Id.* The court effectively gave Dabit leave to file yet another complaint, this time taking care to exclude from the class of plaintiffs anyone who may have purchased, and not merely held, the securities at issue. *Id.* The Second Circuit reasoned that a class limited to holders of securities would give rise to a viable, un-preempted cause of action under state law. *Id.* at 1509.

extend to holders' claims as well as those of purchasers and sellers.¹² In an effort to resolve this circuit split, the U.S. Supreme Court granted certiorari to petitioner Merrill Lynch and held: vacated and remanded.¹³ SLUSA's preemptive scope extends to state-law class actions brought by holders of securities alleging fraudulent manipulation of stock prices.¹⁴

II. RELATED LAW

The stock market crash of 1929 served as the catalyst for the body of federal securities law that continues to regulate securities markets today.¹⁵ Dissatisfied with the ineffective regulatory efforts of the states in this area, Congress passed two major acts calculated to forestall another stock market catastrophe: the Securities Act of 1933 (Securities Act) and the Securities Exchange Act of 1934 (Exchange Act).¹⁶ Section 10(b) of the Exchange Act vested the SEC with the power to enact rules and regulations designed to prevent the use of fraudulent or deceptive devices "in connection with the purchase or sale of any security."¹⁷ Congress expressly autho-

¹² *Kircher v. Putnam Funds Trust*, 403 F.3d 478 (7th Cir. 2005), *vacated*, 126 S. Ct. 2145 (2006).

¹³ *Dabit*, 126 S. Ct. at 1515. *Dabit* was decided 8-0 with Justice Alito not participating. *Id.* at 1506. Justice Stevens authored the unanimous opinion of the Court. *Id.*

¹⁴ *Id.*

¹⁵ Brendan J. McCarthy, "In Connection With": The Need for Limitation to SEC Rule 10b-5 in Dissemination of Misleading Information Cases, 54 CASE W. RES. L. REV. 1347, 1351 (2004).

¹⁶ *Id.*; Joshua D. Ratner, *Stockholders' Holding Claim Class Actions Under State Law After the Uniform Standards Act of 1998*, 68 U. CHI. L. REV. 1035, 1042-43 (2001). The Securities Act dealt primarily with disclosure and fraud in the context of initial public offerings while the Exchange Act focused on the regulation of post-distribution trading. *Blue Chip Stamps v. Manor Drug Stores*, 421 U.S. 723, 752 (1975).

¹⁷ 15 U.S.C. § 78j(b) (2000). The current version of § 78j(b) renders it unlawful:

To use or employ, in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered . . . any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the

rized the SEC to enforce any rules it promulgated under this section.¹⁸ In response to this congressional mandate, the SEC promulgated Rule 10b-5, which roundly outlawed fraud and deceit "in connection with the purchase or sale of any security."¹⁹ The legislative history surrounding § 10(b) is limited.²⁰ Nevertheless, sufficient evidence exists to conclude that Congress intended the section to serve as a "catchall" provision empowering the SEC to crack down on novel deceptive devices jeopardizing the integrity of national securities markets.²¹ In keeping with its legislative pedigree, Rule 10b-5 would ultimately emerge as the chief enforcement mechanism in the field of federal securities law.²²

A. An Implied Private Right of Action Under Rule 10b-5

In drafting § 10(b), Congress did not expressly delineate a private right of action for violations of § 10(b) or its administrative derivation, Rule 10b-5.²³ The U.S. District Court for

protection of investors.¹⁷

15 U.S.C. § 78j(b) (2000).

¹⁸ *Dabit*, 126 S. Ct. at 1509. The Exchange Act created the SEC, providing for the appointment of five commissioners by the President to be confirmed by the Senate. 15 U.S.C. § 78d(a) (2000).

¹⁹ 17 C.F.R. § 240.10b-5(c) (2006). Rule 10b-5 provides that:

It shall be unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce, or of the mails or of any facility of any national securities exchange,

(a) To employ any device, scheme, or artifice to defraud,

(b) To make any untrue statement of a material fact or to omit to state a material fact necessary in order to make the statements made, in the light of the circumstances under which they were made, not misleading, or

(c) To engage in any act, practice, or course of business which operates or would operate as a fraud or deceit upon any person, *in connection with the purchase or sale of any security*.

17 C.F.R. § 240.10b-5 (2006) (emphasis added).

²⁰ McCarthy, *supra* note 15, at 1351.

²¹ *Id.* at 1351-52 (quoting *Ernst & Ernst v. Hochfelder*, 425 U.S. 185, 203 (1976)).

²² *Id.* at 1347.

²³ *Kardon v. Nat'l Gypsum Co.*, 69 F. Supp. 512, 513 (E.D. Penn. 1946); *see also Dabit*, 126 S. Ct. at 1509.

the Eastern District of Pennsylvania was the first court to imply a private right of action in this context in *Kardon v. National Gypsum Co.*²⁴ Some twenty-five years after *Kardon*, the Supreme Court endorsed its prior holding in *Superintendent of Insurance v. Bankers Life & Casualty Co.*²⁵ In *Bankers Life*, a handful of crooked investors, after unscrupulously acquiring all the stock in an insurance company, fraudulently liquidated the company's own bond portfolio to finance the company's takeover.²⁶ The Court granted certiorari to determine whether the insurance company, as a private party, was authorized to sue for alleged violations of Rule 10b-5.²⁷ Noting that the company sustained damages resulting from deceptive practices "touching" its sale of bond securities, the Court acknowledged precedential support for a private right of action under Rule 10b-5.²⁸ In an opinion offering very little reasoning, the Court signaled its affirmation of the already well-entrenched judicially implied private right of action under Rule 10b-5.²⁹

²⁴ See generally *Kardon*, 69 F. Supp. at 514. Judge Kirkpatrick noted that "in view of the general purpose of the [Exchange Act], the mere omission of an express provision for civil liability is not sufficient to negative what the general law implies." *Id.*

²⁵ See *Superintendent of Ins. v. Bankers Life & Cas. Co.*, 404 U.S. 6, 12-13 (1971).

²⁶ *Id.* at 7-8. The swindlers in *Bankers Life* obtained an unsupported check for \$5 million from a co-conspirator. *Id.* Using this check (with no funds behind it), the conspirators bought all the stock of the target company, subsequently installing one of the co-conspirators as president of that company. *Id.* at 8. Using the office of president, the conspirators effected the sale of the company's Treasury bonds, deceiving the company's board into believing that the bond proceeds would be invested in a certificate of deposit. *Id.* Coupling the proceeds of this sale with the cash on hand in the target company, the conspirators acquired sufficient funds to support the phony check they had used to purchase the company in the first place. *Id.*

²⁷ See *id.* at 7, 12-14.

²⁸ *Id.* at 12-13. Justice Douglas, writing for the Court, concluded that "[t]he crux of the present case is that [the plaintiff] suffered an injury as a result of deceptive practices touching its sale of securities as an investor." *Id.* (emphasis added).

²⁹ *Id.* at 13-14.

B. Limiting the Private Right of Action: Standing To Sue Under Rule 10b-5

Several years after recognizing a private right of action under Rule 10b-5 in *Bankers Life*, the Court defined the scope of the right in *Blue Chip Stamps v. Manor Drug Stores*.³⁰ As part of a consent decree, the defendant securities issuer was ordered to offer substantial shares of common stock to its customers, none of which were stockholders prior to the defendant's reorganization.³¹ In compliance with the decree, the defendant distributed a prospectus to all offerees.³² A class consisting of offerees subsequently sued, alleging that the defendant intentionally presented an "overly pessimistic appraisal" of the company in its prospectus.³³ The plaintiffs claimed the defendant issued the negative prospectus to dissuade them from purchasing stock so that the stock could later be offered to the public at a higher price per share.³⁴

The Court granted certiorari to decide whether a cause of action may lie under Rule 10b-5 when the prospective plaintiff has neither purchased nor sold any securities.³⁵ Favorably acknowledging the 'standing rule' developed after *Birnbaum v. Newport Steel Corp.*, which limited the availability of the right of action under Rule 10b-5 to purchasers and sellers of securities, the Court proceeded through a supportive statutory construction analysis before examining key policy considerations militating in favor of the standing rule.³⁶ The Court reasoned

³⁰ *Blue Chip Stamps v. Manor Drug Stores*, 421 U.S. 723 (1975).

³¹ *Id.* at 726. A decade prior to the Court's decision in *Blue Chip Stamps*, the defendant, a trading stamp supplier, had faced reorganization as part of a consent decree stemming from an antitrust action. *Id.* at 725-26. The defendant's customers consisted of stamp retailers. *Id.*

³² *Id.* at 726.

³³ *Id.*

³⁴ *Id.* at 726-27.

³⁵ *Id.* at 727.

³⁶ *Id.* at 731, 733-44; see *Birnbaum v. Newport Steel Corp.*, 193 F.2d 461, 464 (2d Cir. 1952) (narrowly construing the language of § 10(b) and Rule 10b-5 as extending the implied private right of action under Rule 10b-5 to *purchasers and sellers of securities only*). The *Blue Chip Stamps* Court acknowledged that "virtually all lower federal courts facing the issue [of standing under Rule 10b-5] in the

that the peculiar brand of "vexatious litigation" attending securities fraud actions necessitated restricting the scope of action to verifiable purchasers and sellers.³⁷ In addition to noting the dangerous potential for "strike suits" and abusive discovery practices absent such a standing limitation, the Court expressed concern relative to the ease with which non-purchaser/non-seller plaintiffs could advance claims based solely on uncorroborated oral testimony.³⁸ The Court recognized that such claims would generate a settlement value entirely disproportionate to any prospect that the claims would prevail on the merits.³⁹ As a result of these considerations, the Court in *Blue Chip Stamps* restricted the scope of

hundreds of reported cases" decided in the previous twenty-five years had reaffirmed *Birnbaum's* limitation of the cause of action to purchasers and sellers of securities. *Blue Chip Stamps*, 421 U.S. at 731. In expounding upon statutory construction arguments supporting the *Birnbaum* rule, the Court noted that when Congress expressly authorized private civil remedies in the Securities Act and the Exchange Act, it explicitly limited such remedies to purchasers and sellers of securities. *Id.* at 735-36.

³⁷ *Blue Chip Stamps*, 421 U.S. at 739-40. The Court grounded its fears of "vexatious litigation" in two undesirable contingencies attending lawsuits brought by non-purchasers/non-sellers of securities: (1) the prospect that such suits would unreasonably frustrate a defendant's normal business activity given their disproportionately high settlement value and (2) the reliance of such suits on "hazy issues of . . . fact[,] the proof of which [would] depend[] almost entirely on oral testimony." *Id.* at 740, 743 (emphasis added).

³⁸ *Id.* at 740-43. The Court emphasized the significant role of oral testimony in securities fraud actions. *Id.* at 743. Savvy to this fact, the Court recognized that a non-purchaser/non-seller plaintiff could get his case to the jury solely on his uncorroborated claim that he once viewed a prospectus, and the representations contained therein subsequently caused him damage. *Id.* at 746. The Court noted the futility of attempting to defend against such testimony. *Id.* The "virtue of the *Birnbaum* rule," the Court reasoned, is that it restricts the availability of the cause of action to parties who have actually engaged in a verifiable trade, thus injecting "an objectively demonstrable fact" into an otherwise murky evidentiary dynamic. *Id.* at 747.

³⁹ *Id.* at 740-43. The Court reasoned that in the absence of the *Birnbaum* rule, the settlement value of suits brought by non-purchasers/non-sellers would derive in part from the threat of "extensive discovery and disruption of normal business activities" in instances in which the lawsuit might be "groundless in any event." *Id.* at 742-43. The Court concluded that the purchaser-seller limitation espoused in *Birnbaum* would at least limit the class of plaintiffs to parties whose claims might in fact be provable, thus potentially eliminating many frivolous suits. *Id.* at 743, 747.

the private right of action under Rule 10b-5 to purchasers and sellers of securities.⁴⁰

C. Congress Strikes Back at Strike Suits: The Reform Act and SLUSA

Congress passed The Reform Act in 1995 in an effort to alleviate corporate exposure to strike suits in federal court.⁴¹ To encourage voluntary disclosure of forward-looking statements, the Reform Act amended the Exchange Act to create a "safe-harbor" provision limiting liability for investors' reliance on such statements.⁴² Moreover, the Act established heightened pleading standards for private actions brought under Rule 10b-5.⁴³ In response to the tougher federal pleading standards, plaintiffs' attorneys simply abandoned the federal forum and began filing securities fraud actions in state courts.⁴⁴ Since the Reform Act's safe-harbor provisions did not apply to actions filed in state court, this shifting of plaintiffs' claims from federal court to state court effectively undermined the express purposes of the Reform Act.⁴⁵ In an effort to close this frustrating loophole, Congress in 1998 en-

⁴⁰ *Id.* at 755.

⁴¹ Ratner, *supra* note 16, at 1035. The Reform Act is officially known as the Private Securities Litigation Reform Act of 1995. *Id.* Strike suits are suits filed in the hopes that a company will settle—withstanding the merits of the suit—rather than face costly discovery proceedings and the potential for a disproportionately large judgment. *Id.*

⁴² McCarthy, *supra* note 15, at 1358-59. Among other things, the safe-harbor provision requires a plaintiff suing under Rule 10b-5 to prove that the corporate issuer distributed the statement with actual knowledge of its falsity or misleading nature. *Id.* at 1359.

⁴³ *Id.* These heightened pleading standards require plaintiffs suing under Rule 10b-5 to allege claims with greater specificity as to the misleading nature of the defendant's representations and the defendant's state of mind. *Id.* The Reform Act's heightened pleading standards are codified at 15 U.S.C. §§ 77z-1, 78u-4 (2000).

⁴⁴ Ratner, *supra* note 16, at 1047. Studies indicated a marked increase in the volume of class actions filed exclusively in state courts following the Reform Act's implementation. *Id.* Such studies postulated that plaintiffs' attorneys were filing less viable claims in state court in an effort to circumvent the more demanding federal pleading standards. *Id.*

⁴⁵ McCarthy, *supra* note 15, at 1359-60.

acted SLUSA.⁴⁶ SLUSA's core provision precluded private parties from filing covered class actions based on state law in either state or federal court when such actions alleged fraud "in connection with the purchase or sale" of covered securities.⁴⁷ In preempting state-law class actions of this nature, Congress sought to force such actions into federal court where they would be subject to the more stringent federal standards set out in the Reform Act.⁴⁸

D. Defining "In Connection With"

The right of action established and its availability limited, the Court's next challenge would lie in defining the reach of Rule 10b-5. In two relatively recent cases, the Court broadly construed the "in connection with" language of the Rule, thereby greatly expanding the range of conduct that would fall within its proscriptive reach.⁴⁹

In *United States v. O'Hagan*, the Court reviewed the case of an attorney who misappropriated confidential information from a client relative to its plans to initiate the hostile takeover of another entity.⁵⁰ The Court grappled with the issue of whether a non-insider violates Rule 10b-5 when he breaches a fiduciary duty by misappropriating inside information and using it to his advantage in a securities transaction.⁵¹ Chiefly on policy grounds, the Court held that the "in connection with" requirement is satisfied as long as the breach of duty "coincide[s]" with the securities transaction.⁵² The Court reasoned

⁴⁶ Ratner, *supra* note 16, at 1048, 1051; *see also supra* note 1 and accompanying text.

⁴⁷ *See supra* note 2 and accompanying text.

⁴⁸ McCarthy, *supra* note 15, at 1360; *see also supra* note 1.

⁴⁹ *See infra* notes 52, 58.

⁵⁰ *United States v. O'Hagan*, 521 U.S. 642, 647-48 (1997). Anticipating that the entity's stock price would rise upon announcement of the takeover, the attorney purchased common stock and call options in the takeover target. *Id.* He later sold these at a healthy profit when the takeover bid was made public. *Id.* at 648.

⁵¹ *See id.* at 653, 655-56.

⁵² *Id.* at 655-56; *see* Mark J. Loewenstein, *The Supreme Court, Rule 10b-5 and the Federalization of Corporate Law*, 39 IND. L. REV. 17, 27-31 (2005) (observing that the *O'Hagan* Court "push[ed] the 'in connection with' requirement to its out-

that the "misappropriation theory" of liability comported with the Exchange Act's goal of ensuring investor confidence in honest securities markets.⁵³

In *SEC v. Zandford*, the Court reiterated its broad interpretation of "in connection with."⁵⁴ In *Zandford*, an attorney charged with the management of an elderly client's mutual fund wrote checks to himself from the fund, liquidating its stocks to support the checks.⁵⁵ The Court considered the issue of whether the breach of a fiduciary duty of this nature satisfied the "in connection with" requirement of Rule 10b-5.⁵⁶ The Court held that the unauthorized sale of a client's securities for the undisclosed purpose of misappropriating the proceeds does in fact constitute a fraud "in connection with the purchase or sale" of securities.⁵⁷ Echoing the rationale it expounded in *O'Hagan*, the Court reasoned that the breaches here likewise "coincide[d]" with the sale of the client's securities.⁵⁸ Once again, the Court broadly construed "in connection with."⁵⁹

III. *MERRILL LYNCH, PIERCE, FENNER & SMITH, INC. v. DABIT*

In *Dabit*, the Court held that SLUSA preempts state-law class action claims brought by securities holders alleging fraudulent manipulation of stock prices.⁶⁰ More broadly, the

er limit, or perhaps beyond") (quoting 17 C.F.R. § 240.10b-5(c) (2006)).

⁵³ *O'Hagan*, 521 U.S. at 658.

⁵⁴ *SEC v. Zandford*, 535 U.S. 813, 819, 825 (2002).

⁵⁵ *Id.* at 815.

⁵⁶ *See id.* at 818.

⁵⁷ *Id.* at 825.

⁵⁸ *Id.*; *see* Loewenstein, *supra* note 52, at 31-32.

⁵⁹ *See* Loewenstein, *supra* note 52, at 32.

⁶⁰ *Merrill Lynch, Pierce, Fenner & Smith, Inc. v. Dabit*, 126 S. Ct. 1503, 1515 (2006). Following the enactment of SLUSA, plaintiffs' attorneys revived class action "holding claims" in state court as a means of circumventing SLUSA's preemption provision. Ratner, *supra* note 16, at 1036, 1051. A "holding claim" alleges that the defendant misled the plaintiff in such a way that the plaintiff was induced to "hold" his stock rather than sell it. *Id.* at 1036. The damages incurred by the holder ostensibly equate to the decline in the holder's stock value upon a fall in the stock price. *Id.* Prior to *Dabit*, federal courts typically disposed of hold-

Court concluded, as a rule, that SLUSA's preemption provision applies to preempt all state-law class action claims alleging fraud "coincid[ing]" with a securities transaction, regardless of the identity of the plaintiffs.⁶¹

The Court began its analysis by reaffirming its broad construction of SLUSA's pertinent preemption language in the context of earlier precedents.⁶² The Court noted that in adopting the 'standing rule' in *Blue Chip Stamps*, the *Blue Chip Stamps* Court did not purport to define the phrase "in connection with the purchase or sale of [securities]."⁶³ Rather, the "Court purported to define the scope of a private right of action under Rule 10b-5."⁶⁴ The Court noted that on occasions when it *has* endeavored to define the phrase "in connection with the purchase or sale of [securities]," most notably in *Zandford* and *O'Hagan*, it has invariably employed a broad construction requiring only that the alleged fraud "coincide" with a securities transaction.⁶⁵ The Court reasoned that Congress, undoubtedly aware of the Court's broad construction of the phrase in similar contexts, must have intended for this same broad construction to apply to SLUSA when it inserted the identical phrase into SLUSA's core provision.⁶⁶

The Court further reasoned that to restrict the application of SLUSA's preemption provision to purchasers and sellers would undermine the effectiveness of the Reform Act, thus

ing claims by remanding them to state court on the grounds that such claims failed to allege fraud "in connection with the purchase or sale" of a security. *Id.* at 1051. Thus, plaintiffs were able to avoid litigating such actions in federal court, where their claims would be subject to the heightened pleading standards provided for in the Reform Act. *See id.* In this manner, plaintiffs' attorneys effectively "exploit[ed] a loophole" in SLUSA. *Id.* at 1035-36.

⁶¹ *See Dabit*, 126 S. Ct. at 1513, 1515.

⁶² *See id.* at 1512-13.

⁶³ *Id.* at 1512.

⁶⁴ *Id.* Unlike the *Birnbaum* court, which relied on the Rule's text in developing the purchaser-seller limitation, the Court in *Blue Chip Stamps* relied instead on policy considerations in adopting the identical limitation as a "standing rule." *Id.*

⁶⁵ *Id.* at 1513. *See supra* notes 52, 58 and accompanying text.

⁶⁶ *Dabit*, 126 S. Ct. at 1513.

contravening the express purpose of SLUSA.⁶⁷ Recounting its earlier observation that class actions by holders entail a heightened risk of “vexatious litigation,” the Court reasoned that exempting holders from SLUSA’s preemptive application “would be odd, to say the least.”⁶⁸ Moreover, the Court noted that exempting holders from SLUSA’s preemptive coverage would likely create “wasteful, duplicative” litigation.⁶⁹ Conscious both of Congress’s preference for uniform standards and its desire to stem the shift of class actions from federal to state courts, the Court concluded that such unsatisfactory consequences would affront the congressional intent that spawned SLUSA’s enactment.⁷⁰

In reaching its holding, the Court paused to consider the general “presumption that ‘Congress does not cavalierly preempt state-law causes of action.’”⁷¹ The Court noted that in actuality SLUSA does not preempt any state-law causes of action.⁷² Moreover, the Court found Congress’s inclusion in SLUSA of specific exceptions to SLUSA’s preemptive scope to be demonstrative of a “congressional sensitivity to state prerogatives.”⁷³ Consequently, the Court determined that Congress did not act cavalierly in this instance.⁷⁴ As a final consideration, the Court noted that federal law—not state law—has historically served as the primary avenue for filing securities fraud claims as class actions.⁷⁵ Accordingly, the Court reasoned that it was not dealing with a situation in which a federal statute purported to “eliminate[] a historically

⁶⁷ *Id.*

⁶⁸ *Id.* at 1514.

⁶⁹ *Id.* The Court contemplated the prospect of parallel class actions proceeding in state and federal courts, asserting claims arising from identical facts yet governed by different standards. *Id.*

⁷⁰ *Id.*

⁷¹ *Id.* (quoting *Medtronic, Inc. v. Lohr*, 518 U.S. 470, 485 (1996)).

⁷² *Id.* Rather, it merely forecloses plaintiffs’ ability to use the class action device as a vehicle for asserting certain types of claims. *Id.* Both individual plaintiffs and non-covered classes retain the right to pursue any state-law causes of action that may apply. *Id.*

⁷³ *Id.*

⁷⁴ *Id.*

⁷⁵ *Id.*

entrenched state-law remedy."⁷⁶

IV. DISCUSSION

Dabit should be viewed as a judicial preemptive strike against the revival of state-law holding class actions in the wake of SLUSA.⁷⁷ Cognizant of the persisting vulnerability of the Reform Act to adaptive plaintiffs' tactics, the *Dabit* Court sought to foreclose plaintiffs' ability to circumvent the Act by filing holding class actions in state court.⁷⁸ Viewed in this light, *Dabit* effectively closes a loophole that SLUSA failed to close.⁷⁹ Given the implications for the Reform Act absent preemption of such holding claims, the Court's holding in *Dabit* is decidedly defensible on policy grounds alone.⁸⁰ Nevertheless, in addition to its strong policy justifications, *Dabit* comports with the Court's precedents broadly defining the scope of Rule 10b-5's proscriptive reach.⁸¹ Had the Court held otherwise, *Dabit* would oddly conflict with the Court's prior holdings interpreting the meaning of "in connection with."⁸²

At first blush, *Dabit* appears to represent a 'win' for corporate America and, conversely, an erosion of shareholders' legal rights. This perception is an inaccurate one, however. The Court's deference to broad congressional policy objectives evinces its desire to further a primary aim of the Exchange Act: to promote investor confidence by ensuring honesty in securities markets.⁸³ In subjecting state-law class action holding claims to SLUSA's preemptive sweep, the Court indirectly sought to ensure that corporate issuers were not deterred from availing themselves of the safe-harbor provisions outlined in the Reform Act.⁸⁴ By eliminating this new threat of state-law

⁷⁶ *Id.* at 1515.

⁷⁷ See *supra* note 60 and accompanying text.

⁷⁸ See *supra* notes 60, 67-70 and accompanying text.

⁷⁹ See *supra* note 60 and accompanying text.

⁸⁰ See *supra* notes 67-70 and accompanying text.

⁸¹ See *supra* notes 52, 58-59, 62, 65-66 and accompanying text.

⁸² See *supra* notes 52, 58-59 and accompanying text.

⁸³ See *supra* notes 16, 53 and accompanying text.

⁸⁴ See *supra* notes 42, 45, 48 and accompanying text.

liability, the Court's holding will effectively encourage corporate issuers to make proper disclosure of forward-looking statements for the benefit of the investing public.⁸⁵ In a post-Enron, post-WorldCom securities marketplace, the significance of such an advantageous outcome cannot be overstated.

V. CONCLUSION

In *Merrill Lynch, Pierce, Fenner & Smith, Inc. v. Dabit*, the Court found SLUSA's preemptive sweep sufficiently expansive to encompass covered state-law class actions brought by holders of covered securities. In holding that such actions cannot be filed in state court, the Court sought to corral all covered class actions alleging Rule 10b-5 violations into federal court, where such claims would be subject to the heightened federal pleading standards set out in the Reform Act. Given the Court's precedents restricting private actions under Rule 10b-5 to purchasers and sellers of securities, *Dabit* effectively eliminates covered state-law holding class actions from the judicial landscape altogether.

Rex M. Shannon III

⁸⁵ See *supra* note 42 and accompanying text.

Mitchell McNutt & Sams

Attorneys at Law

www.mitchellmcnutt.com

Raise the bar.

**Bradley Arant Rose & White LLP is pleased to announce
that 78 of our partners are recognized in the
2007 edition of *The Best Lawyers in America*.**

We call that raising the bar.


Bradley Arant
BRADLEY ARANT ROSE & WHITE, LLP

BIRMINGHAM CHARLOTTE HUNTSVILLE JACKSON MONTGOMERY WASHINGTON, DC www.bradleyarant.com



Of the nation's Top 100 law firms*,
how many call Mississippi home?

One.

BUTLER SNOW

BUTLER, SNOW, OMARA, STEVENS & CANNADA, PLLC

ATTORNEYS AT LAW

Jackson, MS 601-948-5711 Gulfport, MS 228-864-1170 Memphis, TN 901-680-7200

butlersnow.com

* Based on the BTI Client Relationship Scorecard, a survey of legal decision-makers at Fortune 1000™ companies conducted by BTI Consulting, Wellesley, MA.

FREE BACKGROUND INFORMATION AVAILABLE UPON REQUEST.

**FORMAN
PERRY
WATKINS
KRUTZ &
TARDY LLP**

ATTORNEYS AT LAW

Dallas • Houston • Jackson • New Orleans

Give yourself the advantage of experience

When you need business valuation, forensic accounting or litigation support services, look carefully for professionals who have the credentials, expertise and — most important — the experience required to deliver timely, accurate and reliable service.

At Grantham Poole CPAs, we have what you're looking for. Over the past 30 years, our firm has built a strong reputation for providing solutions that make our clients' lives easier and more profitable. We offer first-class business valuation and litigation support services for a broad range of purposes, including:

- ✓ Shareholder disputes
- ✓ Estate planning and settlement
- ✓ Buy/sell agreements
- ✓ Marital dissolutions
- ✓ Business interruptions
- ✓ Calculation of lost profits
- ✓ Determination of actual and punitive damages
- ✓ Fraud investigations
- ✓ Courtroom testimony
- ✓ Mediation and arbitration
- ✓ Financial data analysis
- ✓ Business sales, mergers and acquisitions

Our highly skilled and experienced professionals hold a variety of relevant credentials, including Certified Public Accountant (CPA), Accredited in Business Valuation (ABV), Certified Valuation Analyst (CVA), Certified Financial Planner (CFP) and Certified Fraud Examiner (CFE).

Everyone in our firm is dedicated to serving our clients and our community with trust, integrity and excellence. We don't just respond to your requests. We proactively anticipate your needs and suggest solutions that give you an advantage, both in and out of court.

Our skilled valuation and litigation support team is eager to help you achieve your objectives:

Stephanie B. Smith, CPA

Robert A. Cunningham, CPA, CFE, CVA

Charles P. Rafferty, CPA, CFE, CVA

J. Thomas Grantham, Jr., CPA, CFP, ABV

George W. Butler, Jr., CPA

James E. Poole, CPA

R. Vance Randall, CPA

We welcome the opportunity to discuss your needs and put our knowledge and experience to work for you. Please call us at 601-957-5050 and let us know how we can be of assistance. ♦

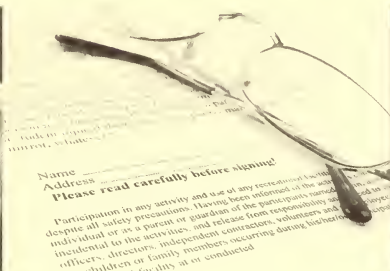
Grantham, Poole, Randall, Reitano, Arrington, & Cunningham, PLLC

Certified Public Accountants

Trusted advisors, solving your problems

MISSISSIPPI BRIEFING SERVICE

...an indispensable service provided by the... **MISSISSIPPI LAW JOURNAL**



*In the legal world, timely and accurate
communication is a key to success...*

How can you stay informed of the latest appellate
decisions without sacrificing your valuable time?

Answer: the Mississippi Briefing Service

Each week, *Recent Decisions of the Mississippi Supreme Court and Court of Appeals* briefs the latest opinions of the Courts. In addition to concise and accurate summaries, each case is supplemented with appropriate and accurate headnotes—which means you find the cases you need and only the cases you need.

Annual subscription to the Mississippi Briefing Service is available for \$85 and is provided in searchable, electronic format via e-mail, providing you the information you need even faster.

To subscribe, or for a free sample brief:

- e-mail:
- write:

MSBRIEF@OLEMISS.EDU
MISSISSIPPI LAW JOURNAL
P.O. Box 849
University, Mississippi 38677



MISSISSIPPI LAW JOURNAL

student-operated, student-lead, practitioner-focused

WHERE
NATURE
AND BUSINESS
seamlessly
CONVERGE.



WOODLANDS OFFICE PARK
Great Venue. Great Value. Great View.



CONVENIENTLY
LOCATED
just minutes away
FROM BUSINESS
AS USUAL.



WOODLANDS OFFICE PARK
Accessible. Affordable. Attractive.



THE
ONLY THING
BETTER THAN
the view
IS THE SERVICE.



WOODLANDS OFFICE PARK
Nearby. Nicely priced. Naturally.



WOODLANDS OFFICE PARK

R I D G E L A N D , M I S S I S S I P P I

For leasing information on the Woodlands contact:

Woodlands Realty Group, LLC

Russell Wilcox, *Broker*

(601) 956-9833

www.woodlandsrealtygroup.com • E-mail: rwilcox@woodlandsrealtygroup.com